

Article

Financial Big Data Solutions for State Space Panel Regression in Interest Rate Dynamics

Dorota Toczydlowska ^{1,*}, Gareth W. Peters ^{2,1,3,4}

¹ Department of Statistical Science, University College London, 1-19 Torrington Place, London WC1E 7HB,

² Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Colin Maclaurin Building, Heriot-Watt University, Edinburgh, EH14 4AS, UK; garethpeters78@gmail.com

³ Man Institute of Quantitative Finance, University of Oxford, Oxford OX1 3BD, UK;

⁴ Systemic Risk Center, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK;

* Correspondence: dtoczydlowska@gmail.com

Received: 12 February 2018; Accepted: 7 July 2018; Published: 17 July 2018

Overview of Appendices

In the following section we provide supplementary information to the studies discussed in the main part of the manuscript,

- in Appendix A we provide the detailed overview of the time series analysed in this study and discussed in Section 2 along with their Bloomberg identifiers;
- in Appendix C we describe the EM algorithm for the standard Gaussian PPCA and its robust variant, robust Gaussian PPCA, as a complementary theory to Section 4;
- in Appendix D we show the proofs to the theorems related to the steps of EM algorithm and discussed in Section 4. The section is divided into the part corresponding to the EM algorithms for t-Student IND and t-Student IID PPCA, respectively.
- in Appendix E we illustrate the results of the synthetic data studies which we conducted to examine the sensitivity, convergence and robustness of various PPCA frameworks. We study the behaviour of the methodologies under various data characteristics such as dimensionality, sample size, proportions of missingness and perturbation and types of the data corruption;
- in Appendix F we detail the supplementary results to the real data case studies.

Appendix A. Data Description

The following section provides detail discussion on financial and macroeconomic data sets used in this study and briefly overviewed in Section 2.

A.1. Euro Libor Yield Curve

The Libor rate is the benchmark of interbank interest rates estimated by large global banks in London. It represents an average cost of short-term loans which are charged when one bank borrows funds from other banks for a given period of time. It is calculated by Intercontinental Exchange and published daily by Thomson Reuters. The currently active maturities for which the Libor rates are specified include 1 day, 1 week, 1 month, 2 months, 3 months, 6 months and 12 months. For further details of the calculation methodology please refer to the ICE Libor official website.

Euro Libor denotes Libor rate for interbank deals denominated in Euro currency. The yield curve used in this study is sourced from Bloomberg using the identifiers given in Table S1.

Table S1. The availability and the Bloomberg identifiers (tickers) of Euro Libor yield curves sourced from Bloomberg.

| Maturity | Ticker | Availability | Maturity | Ticker | Availability |
|----------|---------|-------------------|----------|----------|-------------------|
| 1W | EUR001W | 01/2006 - 08/2017 | 7Y | EUSW7V3 | 01/2006 - 08/2017 |
| 1M | EUR001M | 01/2006 - 08/2017 | 8Y | EUSW8V3 | 01/2006 - 08/2017 |
| 2M | EUR002M | 01/2006 - 08/2017 | 9Y | EUSW9V3 | 01/2006 - 08/2017 |
| 3M | EUR003M | 01/2006 - 08/2017 | 10Y | EUSW10V3 | 01/2006 - 08/2017 |
| 1Y | EUSW1VC | 01/2006 - 08/2017 | 15Y | EUSW15V3 | 01/2006 - 08/2017 |
| 2Y | EUSW2V3 | 01/2006 - 08/2017 | 20Y | EUSW20V3 | 01/2006 - 08/2017 |
| 3Y | EUSW3V3 | 01/2006 - 08/2017 | 25Y | EUSW25V3 | 01/2006 - 08/2017 |
| 5Y | EUSW5V3 | 01/2006 - 08/2017 | | | |

The time series span from 01/01/2006 - 14/08/2017 for maturities 1 week (1W), 1 month (1M), 2 months (2M), 3 months (3M), 1 year (1Y), 2 years (2Y), 3 years (3Y), 5 years (5Y), 7 years (7Y), 8 years (8Y), 9 years (9Y), 10 years (10Y), 15 years (15Y), 20 years (20Y) and 25 years (25Y). We use Euro Libor rates provided by Intercontinental Exchange up to 12 months, whereas additional 1Y and longer maturities are represented by swap rates on Euro Libor. The panels in Figure S1 present the time series of the Euro Libor yield curve at fixed tenors over time (left) and indicators of missing values in the data over time (right) denoted by single crosses. The reader may notice that the rates corresponding to the maturities less than 1 year, and provided by Intercontinental Exchange, are less volatile than the instruments constructed using swap rates. In addition, the number of values which are not available over time (are not observed) is much greater for smaller maturities and their pattern is more regular. Furthermore, one can observe that the swap-related rates have the majority of missing values before 2008 and 1Y having none.

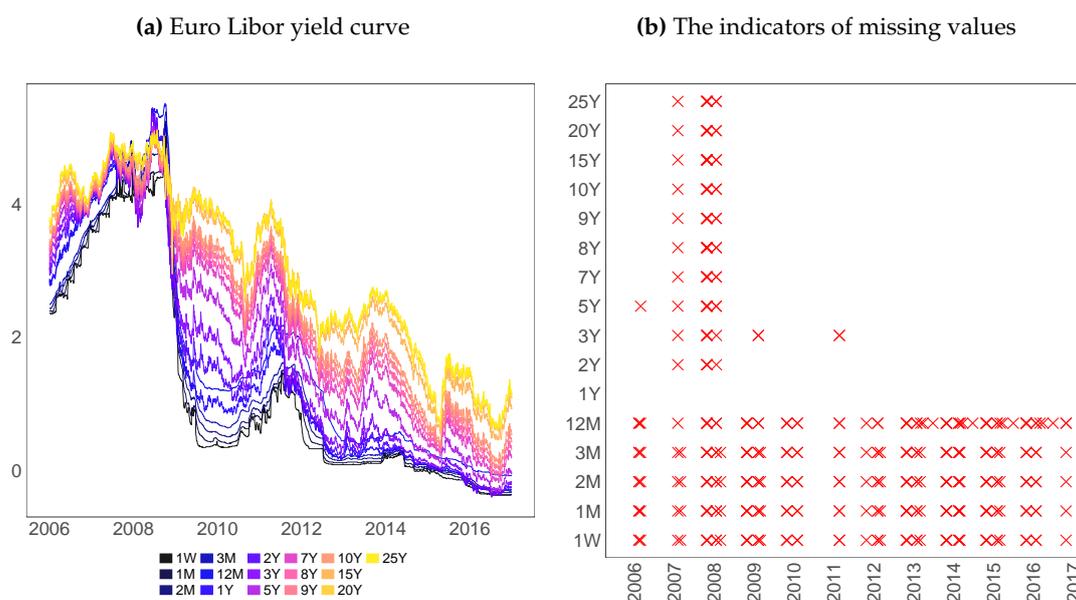


Figure S1. The Euro Libor rates (%) (left panel) and the indicators of missing values (right panel) over time. The different colours of lines in the left panel correspond to the rates with different maturities given in the legend. In the right panel, a single cross indicates daily unavailability of an instrument. The missingness does not correspond to weekends and holidays.

A.2. Country-Specific Sovereign Yield Curves

The full sample of country-specific sovereign yield curves consists of 94 time series of interest rates corresponding to different maturities from the 12 countries listed in Table 1. The examined sample spans 01/2006 - 12/2016. The sovereign yield curves are downloaded from Bloomberg using the search query 'GGR' which lists generic rates for international government securities. The yields are comprised of the most recently issued or closest current nominal maturity government security

over time. One can locate the underlying benchmark bonds under Bloomberg identifiers of sovereign yield curves, given in Table S2.

Table S2. The availability and Bloomberg indentifiers (tickers) of the sovereign yield curves and inflation-linked yield curves sourced from Bloomberg. The column Type indicate what are the benchmark instruments of the inflation-linked yields, bond rates (1) or swap rates (2)

| Country | Maturity | Sovereign Yield Curve | | Inflation-linked Yield Curve | |
|---------|--------------|-----------------------|-------------------|------------------------------|-------------------|
| | | Ticker | Availability | Ticker | Availability |
| AU | 3M | GACGB3M Index | 05/2009 - 12/2016 | | |
| | 1Y | GACGB1 Index | 01/2006 - 12/2016 | AUSWIT1 Index | 06/2007 - 12/2016 |
| | 2Y | GACGB2 Index | 01/2006 - 12/2016 | AUSWIT2 Index | 06/2007 - 12/2016 |
| | 5Y | GACGB5 Index | 01/2006 - 12/2016 | AUSWIT5 Index | 06/2007 - 12/2016 |
| | 7Y | GACGB7 Index | 01/2006 - 12/2016 | AUSWIT7 Index | 06/2007 - 12/2016 |
| | 10Y | GACGB10 Index | 01/2006 - 12/2016 | AUSWIT10 Index | 06/2007 - 12/2016 |
| | 15Y | GACGB15 Index | 01/2006 - 12/2016 | AUSWIT15 Index | 06/2007 - 12/2016 |
| | 20Y | GACGB20 Index | 11/2013 - 12/2016 | AUSWIT20 Index | 06/2007 - 12/2016 |
| | 25Y | | | AUSWIT25 Index | 06/2007 - 12/2016 |
| BZ | 3M | GEBR03M Index | 03/2007 - 12/2016 | | |
| | 6M | GEBR06M Index | 03/2007 - 12/2016 | | |
| | 1Y | GEBR01Y Index | 03/2007 - 12/2016 | | |
| | 2Y | GEBR02Y Index | 03/2007 - 12/2016 | | |
| | 5Y | GEBR5Y Index | 01/2007 - 12/2016 | | |
| | 10Y | GEBR10Y Index | 01/2007 - 12/2016 | | |
| DE | 1Y | GDBR1 Index | 01/2006 - 12/2016 | GRSWIT1 Curncy | 04/2006 - 12/2016 |
| | 2Y | GDBR2 Index | 01/2006 - 12/2016 | GRSWIT2 Curncy | 04/2006 - 12/2016 |
| | 5Y | GDBR5 Index | 01/2006 - 12/2016 | GRSWIT5 Curncy | 04/2006 - 12/2016 |
| | 7Y | GDBR7 Index | 01/2006 - 12/2016 | GEIL5Y Index | 12/2008 - 12/2015 |
| | | GDBR7 Index | 01/2006 - 12/2016 | GRSWIT7 Curncy | 04/2006 - 12/2016 |
| | 10Y | GDBR10 Index | 01/2006 - 12/2016 | GEIL7Y Index | 12/2008 - 12/2016 |
| | | GDBR10 Index | 01/2006 - 12/2016 | GRSWIT10 Curncy | 04/2006 - 12/2016 |
| | 15Y | GDBR15 Index | 05/2008 - 12/2016 | GEIL10Y Index | 06/2009 - 12/2016 |
| | 20Y | GDBR20 Index | 01/2006 - 12/2016 | GRSWIT15 Curncy | 04/2006 - 12/2016 |
| | 25Y | | | GRSWIT20 Curncy | 04/2006 - 12/2016 |
| | | | GRSWIT25 Curncy | 04/2006 - 12/2016 | |
| ES | 3M | GSPG3M Index | 01/2006 - 12/2016 | | |
| | 6M | GSPG6M Index | 01/2006 - 12/2016 | | |
| | 1Y | GSPG12M Index | 01/2006 - 12/2016 | SPSWIT1 Index | 01/2006 - 12/2016 |
| | 2Y | GSPG2YR Index | 01/2006 - 12/2016 | SPSWIT2 Index | 01/2006 - 12/2016 |
| | 5Y | GSPG5YR Index | 01/2006 - 12/2016 | SPSWIT5 Index | 01/2006 - 12/2016 |
| | 7Y | GSPG7YR Index | 01/2006 - 12/2016 | SPSWIT7 Index | 01/2006 - 12/2016 |
| | 10Y | GSPG10YR Index | 01/2006 - 12/2016 | SPSWIT10 Index | 01/2006 - 12/2016 |
| | 15Y | GSPG15YR Index | 01/2006 - 12/2016 | SPSWIT15 Index | 01/2006 - 12/2016 |
| | 20Y | | | SPSWIT20 Index | 01/2006 - 12/2016 |
| | 25Y | | | SPSWIT25 Index | 01/2006 - 12/2016 |
| FR | 3M | GBTF3MO Index | 01/2006 - 12/2016 | | |
| | 6M | GBTF6MO Index | 01/2006 - 12/2016 | | |
| | 1Y | GBTF1YR Index | 01/2006 - 12/2016 | FRSW1 Index | 01/2006 - 12/2016 |
| | 2Y | GFRN2 Index | 01/2006 - 12/2016 | GFRGIN02 Index | 04/2009 - 06/2016 |
| | | GFRN2 Index | 01/2006 - 12/2016 | FRSW12 Index | 01/2006 - 12/2016 |
| | 5Y | GFRN5 Index | 01/2006 - 12/2016 | GFRGIN05 Index | 06/2006 - 12/2016 |
| | | GFRN5 Index | 01/2006 - 12/2016 | FRSW15 Index | 01/2006 - 12/2016 |
| | 7Y | GFRN7 Index | 01/2006 - 12/2016 | FRSW17 Index | 01/2006 - 12/2016 |
| | | GFRN10 Index | 01/2006 - 12/2016 | GFRGIN10 Index | 06/2006 - 12/2016 |
| | 10Y | GFRN10 Index | 01/2006 - 12/2016 | FRSW10 Index | 01/2006 - 12/2016 |
| | | GFRN15 Index | 01/2006 - 12/2016 | GFRGIN15 Index | 04/2009 - 12/2016 |
| | 15Y | GFRN15 Index | 01/2006 - 12/2016 | FRSW15 Index | 01/2006 - 12/2016 |
| | | GFRN20 Index | 01/2006 - 12/2016 | FRSW20 Index | 01/2006 - 12/2016 |
| | 20Y | GFRN20 Index | 01/2006 - 12/2016 | GFRGIN20 Index | 01/2008 - 06/2012 |
| | 25Y | | | FRSW25 Index | 01/2006 - 12/2016 |
| GB | 3M | GUKG3M Index | 12/2008 - 12/2016 | | |
| | 6M | GUKG6M Index | 08/2007 - 11/2016 | | |
| | 1Y | GUKG1 Index | 01/2006 - 12/2016 | BPSWIT1 Index | 01/2006 - 12/2016 |
| | 2Y | GUKG2 Index | 01/2006 - 12/2016 | GUKGIN02 Index | 01/2007 - 02/2010 |
| | | GUKG2 Index | 01/2006 - 12/2016 | BPSWIT2 Index | 01/2006 - 12/2016 |
| | 5Y | GUKG5 Index | 01/2006 - 12/2016 | GUKGIN05 Index | 01/2006 - 12/2016 |
| | | GUKG5 Index | 01/2006 - 12/2016 | BPSWIT5 Index | 01/2006 - 12/2016 |
| | 7Y | GUKG7 Index | 01/2006 - 12/2016 | BPSWIT7 Index | 01/2006 - 12/2016 |
| | | GUKG10 Index | 01/2006 - 12/2016 | GUKGIN10 Index | 01/2006 - 12/2016 |
| | 10Y | GUKG10 Index | 01/2006 - 12/2016 | BPSWIT10 Index | 01/2006 - 12/2016 |
| | | GUKG15 Index | 01/2006 - 12/2016 | BPSWIT15 Index | 01/2006 - 12/2016 |
| | 15Y | GUKG15 Index | 01/2006 - 12/2016 | GUKGIN20 Index | 01/2006 - 12/2016 |
| | | GUKG20 Index | 01/2006 - 12/2016 | BPSWIT20 Index | 01/2006 - 12/2016 |
| 25Y | GUKG20 Index | 01/2006 - 12/2016 | BPSWIT25 Index | 01/2006 - 12/2016 | |
| IR | 3M | GIGB3M Index | 10/2013 - 11/2015 | | |
| | 1Y | GIGB1YR Index | 01/2006 - 12/2016 | | |
| | 2Y | GIGB2YR Index | 01/2006 - 12/2016 | | |
| | 5Y | GIGB5YR Index | 01/2006 - 06/2016 | | |
| | 7Y | GIGB7YR Index | 02/2006 - 12/2016 | | |
| | 15Y | GIGB15YR Index | 01/2006 - 12/2016 | | |
| IT | 3M | GBOTG3M Index | 01/2006 - 12/2016 | | |
| | 6M | GBOTG6M Index | 01/2006 - 12/2016 | | |
| | 1Y | GBOTG12M Index | 01/2006 - 12/2016 | GRITL1 Index | 06/2007 - 12/2016 |
| | | | ILSW11 Index | 01/2006 - 12/2016 | |

Continue on the next page

Table S2. The availability and the reference to the Bloomberg terminal (tickers) of the sovereign yield curves and inflation-linked yield curves sourced from Bloomberg (cont.).

| Country | Maturity | Sovereign Yield Curve | | Inflation-linked Yield Curve | |
|---------|----------------|-----------------------|-------------------|------------------------------|-------------------|
| | | Ticker | Availability | Ticker | Availability |
| | 2Y | GBT2YR Index | 01/2006 - 12/2016 | ILSW12 Index | 01/2006 - 12/2016 |
| | 5Y | GBT5YR Index | 01/2006 - 12/2016 | GRITL5 Index | 06/2007 - 12/2016 |
| | | GBT15YR Index | 01/2006 - 12/2016 | ILSW15 Index | 01/2006 - 12/2016 |
| | 7Y | GBT7YR Index | 01/2006 - 12/2016 | ILSW17 Index | 01/2006 - 12/2016 |
| | | GBT10YR Index | 01/2006 - 12/2016 | GRITL10 Index | 06/2007 - 12/2016 |
| | 10Y | GBT10YR Index | 01/2006 - 12/2016 | ILSW10 Index | 01/2006 - 12/2016 |
| | | GBT15YR Index | 02/2006 - 12/2016 | GRITL15 Index | 05/2009 - 12/2016 |
| | 15Y | GBT15YR Index | 02/2006 - 12/2016 | ILSW15 Index | 01/2006 - 12/2016 |
| | 20Y | GBT20YR Index | 02/2006 - 12/2016 | ILSW20 Index | 01/2006 - 12/2016 |
| | 25Y | | | ILSW25 Index | 01/2006 - 12/2016 |
| JP | 3M | GJTB3MO Index | 01/2006 - 12/2016 | | |
| | 6M | GJTB6MO Index | 01/2006 - 12/2016 | | |
| | 1Y | GJGB1 Index | 01/2006 - 12/2016 | JYSWIT1 Index | 03/2007 - 12/2016 |
| | 2Y | GJGB2 Index | 01/2006 - 12/2016 | JYSWIT2 Index | 03/2007 - 12/2016 |
| | 5Y | GJGB5 Index | 01/2006 - 12/2016 | JYSWIT5 Index | 03/2007 - 12/2016 |
| | | GJGB5 Index | 01/2006 - 12/2016 | GJYGIN05 Index | 07/2009 - 04/2014 |
| | 7Y | GJGB7 Index | 01/2006 - 12/2016 | GJYGIN07 Index | 07/2009 - 07/2012 |
| | | GJGB7 Index | 01/2006 - 12/2016 | JYSWIT7 Index | 03/2007 - 12/2016 |
| | 10Y | GJGB10 Index | 01/2006 - 12/2016 | JYSWIT10 Index | 03/2007 - 12/2016 |
| | | GJGB10 Index | 01/2006 - 12/2016 | GJYGIN10 Index | 01/2006 - 12/2016 |
| | 15Y | GJGB15 Index | 01/2006 - 12/2016 | JYSWIT15 Index | 03/2007 - 12/2016 |
| | 20Y | GJGB20 Index | 01/2006 - 12/2016 | JYSWIT20 Index | 03/2007 - 12/2016 |
| | 25Y | | | JYSWIT25 Index | 03/2007 - 03/2015 |
| PO | 3M | GSPT3M Index | 01/2006 - 12/2016 | | |
| | 6M | GSPT6M Index | 01/2006 - 12/2016 | | |
| | 1Y | GSPT12M Index | 01/2006 - 12/2016 | | |
| | 2Y | GSPT2YR Index | 01/2006 - 12/2016 | | |
| | 5Y | GSPT5YR Index | 01/2006 - 12/2016 | | |
| | 7Y | GSPT7YR Index | 01/2006 - 12/2016 | | |
| 10Y | GSPT10YR Index | 01/2006 - 12/2016 | | | |
| 15Y | GSPT15YR Index | 02/2009 - 12/2016 | | | |
| SA | 1Y | | | SASWIT1 Curncy | 05/2007 - 12/2016 |
| | 2Y | GSAB2YR Index | 01/2006 - 12/2016 | SASWIT2 Curncy | 05/2007 - 12/2016 |
| | 5Y | GSAB5YR Index | 01/2006 - 12/2016 | SASWIT5 Curncy | 05/2007 - 12/2016 |
| | 7Y | GSAB7YR Index | 07/2007 - 12/2016 | SASWIT7 Curncy | 05/2007 - 12/2016 |
| | 10Y | GSAB10YR Index | 01/2006 - 12/2016 | SASWIT10 Curncy | 05/2007 - 12/2016 |
| | 15Y | GSAB15YR Index | 01/2006 - 12/2016 | SASWIT15 Curncy | 05/2007 - 12/2016 |
| | 20Y | GSAB20YR Index | 01/2006 - 12/2016 | SASWIT20 Curncy | 05/2007 - 12/2016 |
| 25Y | | | SASWIT25 Curncy | 05/2007 - 12/2016 | |
| US | 6M | USGG6M Index | 01/2006 - 12/2016 | | |
| | 1Y | USGG12M Index | 06/2008 - 12/2016 | USSWIT1 Index | 01/2006 - 12/2016 |
| | 2Y | USGG2YR Index | 01/2006 - 12/2016 | USSWIT2 Index | 01/2006 - 12/2016 |
| | 5Y | USGG5YR Index | 01/2006 - 12/2016 | USGGT05Y Index | 01/2006 - 12/2016 |
| | | USGG5YR Index | 01/2006 - 12/2016 | USSWIT5 Index | 01/2006 - 12/2016 |
| | 7Y | USGG7YR Index | 02/2009 - 12/2016 | USSWIT7 Index | 01/2006 - 12/2016 |
| | | USGG10YR Index | 01/2006 - 12/2016 | USGGT10Y Index | 01/2006 - 12/2016 |
| | 10Y | USGG10YR Index | 01/2006 - 12/2016 | USSWIT10 Index | 01/2006 - 12/2016 |
| | 15Y | | | USSWIT15 Index | 01/2006 - 12/2016 |
| | 20Y | | | USSWIT20 Index | 01/2006 - 12/2016 |
| 25Y | | | USGGT20Y Index | 01/2006 - 12/2016 | |
| | | | USSWIT25 Index | 01/2006 - 12/2016 | |

The Figure S2 illustrates the evolution of sovereign yield curves (left subfigure) and the associate unavailability of the yield over time (right subfigure), where each panel corresponds to country-specific statistics. The yield curves are characterized by both, irregular sets of available maturities and unevenly spaced time series. The full information related to the availability of the instruments and their Blomberg identifiers are given in Table S2.

The irregular availability of time of the country-specific yield curves illustrated in the right panel of Figure S2 is manifest in two distinct types of patterns highlighted daily for a given instrument by a single red cross (when missing). The first pattern is related to the different dates when instruments have started to be issued, for instance, the Irish sovereign yield curve at 3M maturity only began to be available from 10/2013. Therefore, we face constantly in time unavailability of this instrument which is indicated on the plots as a solid straight red line (constructed from a dense set of red crosses) which starts from the beginning of the sample. The second pattern appears when we have a sparse set of dates for which the yields are only temporally not available. It can be the result of a temporal lack of financial securities liquid enough to calculate the rate. This pattern is highlighted by single, sparse groups of red crosses. For instance, one may notice short unavailabilities of 3M, 6M and 1Y maturities from the French sovereign yield curve. Figure S4 and Figure S5 summarises the overall proportion of

missing values per a sovereign bond rate across different countries and maturities given the sample size of 2 870 days (excluding holidays and weekends).

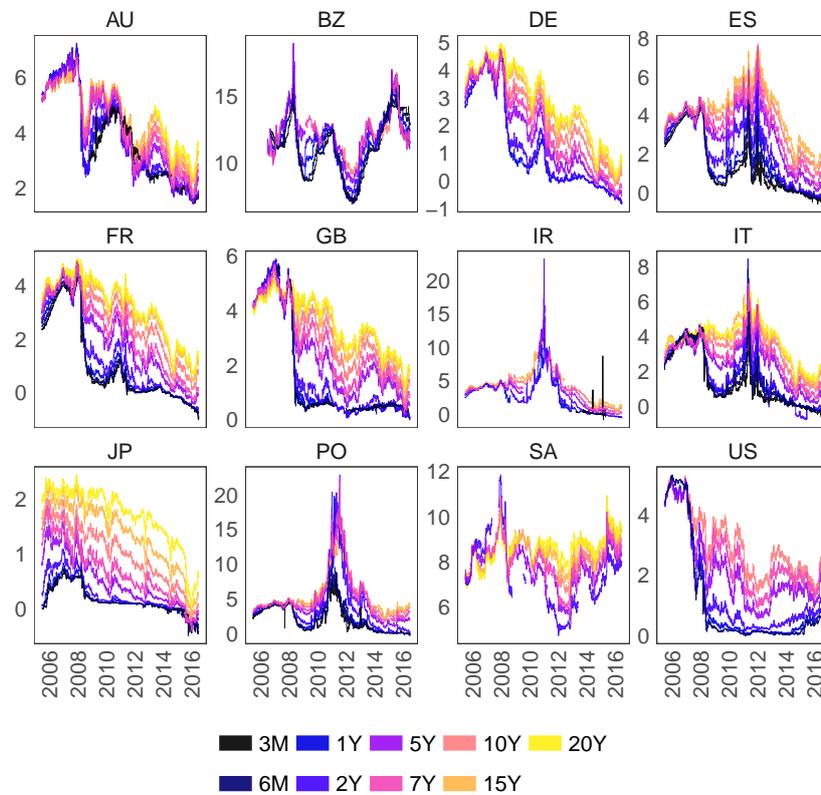


Figure S2. The sovereign yield curves (%) of 12 countries listed in Table 1. The colors of lines correspond to the different maturities.

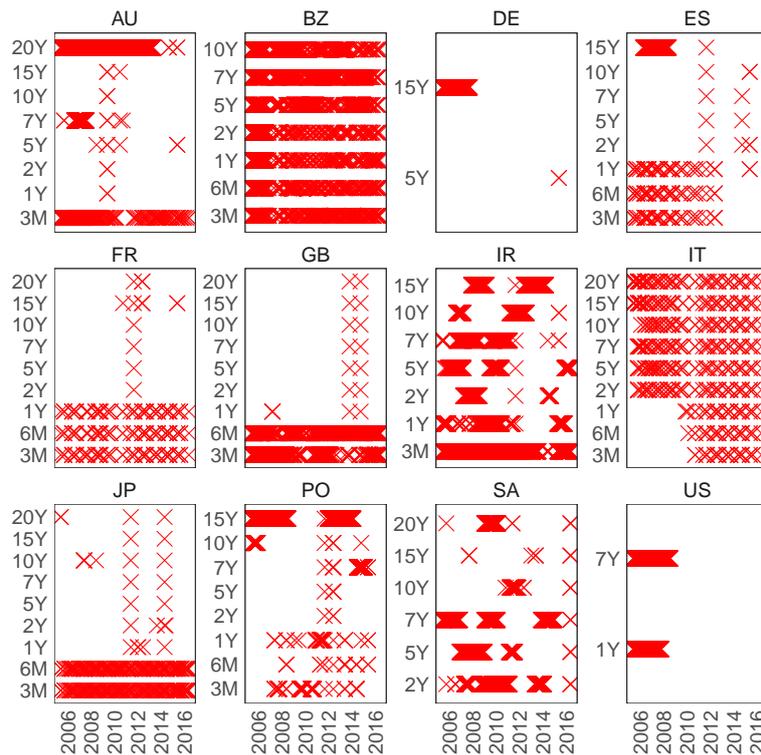


Figure S3. The indicators of missing values for sovereign yields of 12 countries listed in Table 1. The red crosses correspond to daily unavailabilities of rates at various maturities (y-axis) over time (x-axis). The missigness does not correspond to weekends and holidays.

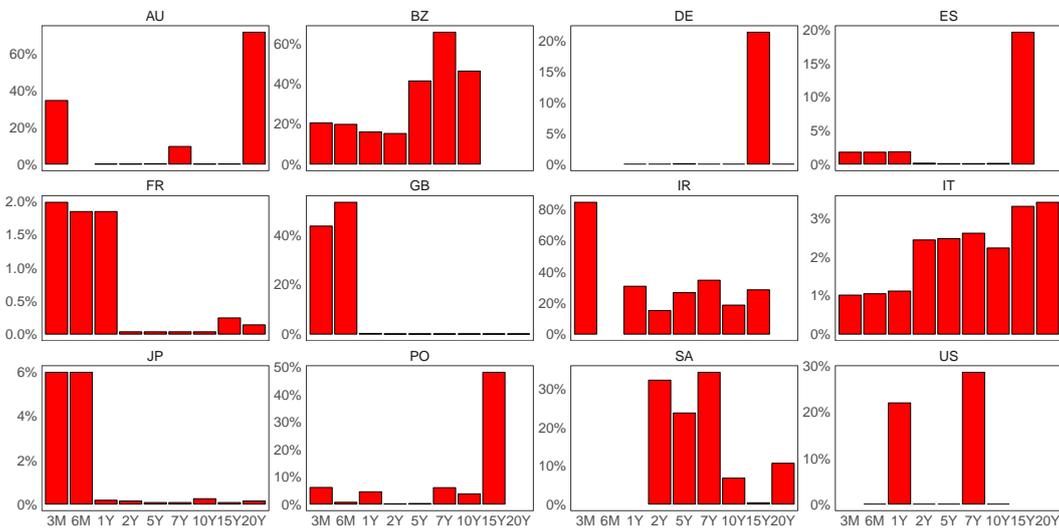


Figure S4. The proportion (%) of missing values present in the time series of sovereign bond rates across different maturities for a sample period 01/2006 - 12/2016 (2870 days excluding weekends and holidays). The blank spaces on the x-axis correspond to instruments which are not available in Bloomberg. The patterns of missingness are illustrated in Figure S3. The missingness does not correspond to weekends and holidays.

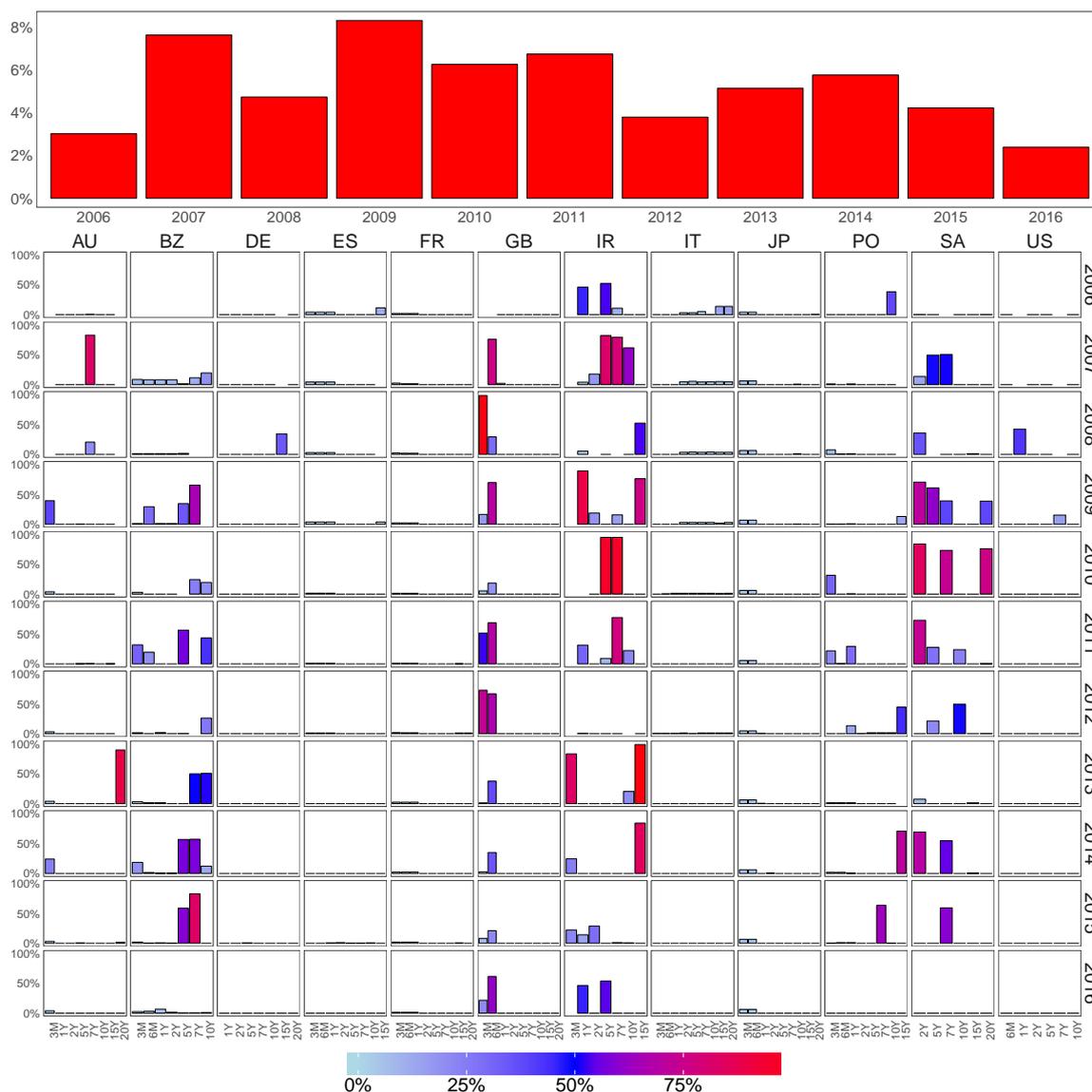


Figure S5. The percentage of the aggregated missing entries per year (top panel) and the percentage of missing entries per year (column-wise) and country (row-wise) of yields across maturities (x-axis) from sovereign yields. The colours of bars in the bottom plot correspond to the proportion of the missingness.

A.3. Measures of Inflation Risk

The data under the category **Inflation** consists of monthly Consumer Price Indexes from 12 countries and 13 daily end-of-the-day inflation-indexed swap or bond based yield curves, from 8 major global economies: Germany, France, Spain, Italy, United Kingdom, Japan, Australia and the United States of America. The country-specific rates are referenced against CPI indexes.

A.3.1. Country-Specific Consumer Price Index

The nation-wide Consumer Price Index (CPI) is usually tailored to the countries' objectives and therefore differs across different economies with regard to the products included into its construction.

In order to ensure that the CPIs among 12 economies are calculated using a methodology which is unified as much as possible across different countries, we use the Harmonised Index of Consumer Prices (HICP) for European Union members which reflects monthly changes. The monthly time series of the index are calculated by Eurostat on behalf of National Statistics Institutes (NSIs) of the European

Union member states and European Free Trade Association (EFTA) which provides the guidelines for calculating HICP, please refer to the official website for details. The term 'harmonised' highlights the fact that all the countries in the European Union follow the same methodology. In the report provided by the Office of National Statistics in Eidukas (2016), the author exemplifies the most frequent differences between CPIs and HICP, i.e. showing that the majority of CPIs excludes the expenditures of non-residents in the economic territory. The HICP and CPI are available from Organisation for Economic Cooperation and Development (OECD) database.

Table S3. The availability and Bloomberg identifiers (tickers) of the national macroeconomic measures of economic activity: Consumer Price Index (CPI), Gross Domestic Product (GDP), Labour Productivity (LP), Unemployment Rates (UE) and German Bund Open Interest, 2W Euro repo rate, 5Y Markit Intraxx Index and exchange rates between major global economies.

| Instrument | Country | Ticker | Availability | Instrument | Country | Ticker | Availability |
|---------------------------|----------------|---------------------------|-------------------|------------------------|-------------------|-----------------|-------------------|
| Consumer Price Index | AU | AUCPIYOY Index | 03/2006 - 12/2016 | Gross Domestic Product | AU | WGDPAUST Index | 12/2006 - 12/2016 |
| | BZ | BZPIPCY Index | 01/2006 - 12/2016 | | BZ | WGDPBRAZ Index | 12/2006 - 12/2016 |
| | DE | ECCPDEMM Index | 01/2006 - 12/2016 | | DE | WGDPGERM Index | 12/2006 - 12/2016 |
| | ES | ECCPESMM Index | 01/2006 - 12/2016 | | ES | WGDPSPAI Index | 12/2006 - 12/2016 |
| | FR | ECCPFRMM Index | 01/2006 - 12/2016 | | FR | WGDPFRAN Index | 12/2006 - 12/2016 |
| | GB | ECCPUKMM Index | 01/2006 - 12/2016 | | GB | WGDPUK Index | 12/2006 - 12/2016 |
| | IR | ECCPIEMM Index | 01/2006 - 12/2016 | | IR | WGDPIREL Index | 12/2006 - 12/2016 |
| | IT | ECCPITMM Index | 01/2006 - 12/2016 | | IT | WGDPITAL Index | 12/2006 - 12/2016 |
| | JP | JNCPIMOM Index | 01/2006 - 12/2016 | | JP | WGDPJAPA Index | 12/2006 - 12/2016 |
| | PO | ECCPPTMM Index | 01/2006 - 12/2016 | | PO | WGDPPOINT Index | 12/2006 - 12/2016 |
| | SA | SACPIYOY Index | 01/2006 - 12/2016 | | SA | WGDPSSOUT Index | 12/2006 - 12/2016 |
| US | CPI YOY Index | 01/2006 - 12/2016 | US | WGDPUS Index | 12/2006 - 12/2016 | | |
| Labour Productivity | AU | OEEOAULP Index | 03/2006 - 12/2016 | Unemployment Rate | AU | EHUPAU Index | 03/2006 - 12/2016 |
| | BZ | OEEOBRLF Index | 12/2006 - 12/2016 | | BZ | EHUPBR Index | 03/2006 - 12/2016 |
| | DE | OEEODELP Index | 03/2006 - 12/2016 | | DE | EHUPDE Index | 12/2005 - 12/2016 |
| | ES | OEEOESLP Index | 12/2006 - 12/2016 | | ES | EHUPES Index | 03/2006 - 12/2016 |
| | FR | OEEOFRLP Index | 03/2006 - 12/2016 | | FR | EHUPFR Index | 03/2006 - 12/2016 |
| | GB | OEEOGBLP Index | 03/2006 - 12/2016 | | GB | EHUPGB Index | 03/2006 - 12/2016 |
| | IR | OEEOIELP Index | 03/2006 - 12/2016 | | IR | EHUPIE Index | 03/2006 - 12/2016 |
| | IT | OEEOITLP Index | 03/2006 - 12/2016 | | IT | EHUPIT Index | 03/2006 - 12/2016 |
| | JP | OEEOJPLP Index | 12/2005 - 12/2016 | | JP | EHUPJP Index | 03/2006 - 12/2016 |
| | PO | OEEOPTLP Index | 03/2006 - 12/2016 | | PO | EHUPPT Index | 03/2006 - 12/2016 |
| | SA | OEEOZALF Index | 12/2005 - 12/2016 | | SA | EHUPZA Index | 03/2006 - 12/2016 |
| US | OEEOUSLP Index | 03/2006 - 12/2016 | US | EHUPUS Index | 03/2006 - 12/2016 | | |
| German Bund Open Interest | | RX1 Comdty | 01/2006 - 12/2016 | | | | |
| 3M Euro repo rate | | EURR003M Index | 01/2006 - 12/2014 | | | | |
| 3M Euribor rate | | EURR003M Index | 01/2006 - 12/2016 | | | | |
| 5Y Markit Intraxx | | ITRX EUR CDSI GEN 5Y Corp | 10/2011 - 12/2016 | | | | |
| FX | AUDUSD | AUDUSD Currency | 01/2006 - 12/2016 | | | | |
| | GBPUSD | GBPUSD Currency | 01/2006 - 12/2016 | | | | |
| | USDJPY | USDJPY Currency | 01/2006 - 12/2016 | | | | |
| | EURUSD | EURUSD Currency | 01/2006 - 12/2016 | | | | |

The CPIs for Japan, South Africa and the United States of America, are specified on a monthly basis, whereas the CPI of Australian is available quarterly and illustrates yearly changes. The indexes are calculated by national statistical institutions: Australian Bureau of Statistics, Ministry of Internal Affairs and Communications, Statistics South Africa and Bureau of Labor Statistics respectively. As a measure of Brazilian Consumer Price Index we use monthly time series of Extended National Consumer Price Index) which covers families living in 11 urban areas of Brazil: Belem, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, Sao Paulo, Curitiba, Porto Alegre, Brasilia and Goiania and whose monthly income is between 1 and 40 times the Brazilian minimum wage. The Bloomberg identifiers of the collected CPIs are listed in Table S3 and the evolution of indexes over time is illustrated in the top left panel of Figure S9.

A.3.2. Country-Specific Inflation-Linked Yield Curves

The collected inflation-linked sovereign yield curves were sourced following similar steps adopted for the sovereign yield curves via a search query of Bloomberg generic yield curves, 'GGR'. The sets of inflation-linked yields are shorter and consists of yield curves for 9 countries as indicated in Table 1. The Bloomberg identifiers of elements of the yields which correspond to a single maturity, and their availabilities are listed in Table S2. The collected set of inflation-linked sovereign yields can be partitioned into two groups, the yields which are comprised of inflation-indexed government securities, and those which are comprised of inflation-linked swap rates. The inflation-linked instruments are referenced against a consumer price index calculated monthly by national statistics institutes.

The most commonly used indicator of inflation expectations is derived as the difference between a nominal conventional and inflation-linked bonds yields at similar maturity. However, with the rapid

growth of the swap market starting from 2003, it is nowadays much easier to construct comprehensive inflation-linked yield curve which is comprised of inflation-linked swap rates, rather than sovereign bonds, due to their broader availability. Various studies of global market regulators such as Schulz and Stapf (2009), investigate the quality of the prediction given by swap-based inflation-linked sovereign yield curve. The authors of Schulz and Stapf (2009) argue that the measures of inflation expectations extracted from inflation linked-swaps bias the indicator of inflation expectations. They indicate the inflation risk premium and liquidity premium which is mitigated from the objectives of the inflation-linked swap market participants. For instance, the corporations, which revenue is linked to inflation, hedge against the risk of low inflation, or pension funds and insurance companies, which hedge against the risk of high inflation and rising prices. The panels in Figure S6 show the evolution of inflation-linked swap (the blue line) and sovereign bonds (the black line) yield curves across different maturities (column-wise) for 8 countries. The corresponding country-specific process of missing values indicator is illustrated in left and right panels of Figures S8, respectively, with proportions of missing values per the whole sample listed in Figure S4.

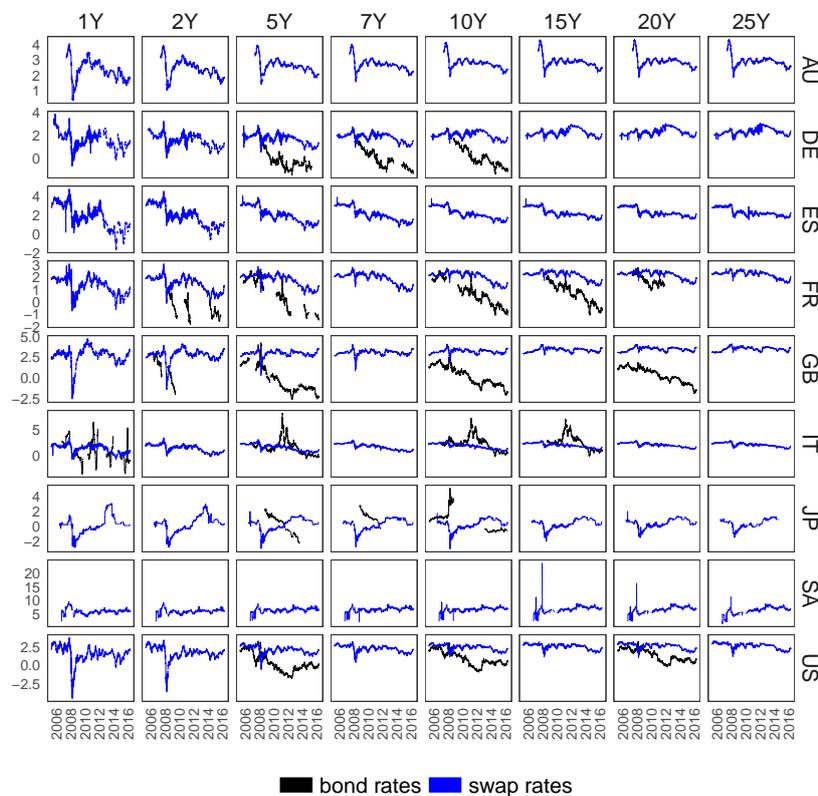


Figure S6. The comparison of inflation-lined yield curves (%) obtained from inflation-linked government bonds (black lines) or inflation-linked swap rates (blue lines) of 8 countries and Euro- zone listed in Table 1. The column-wise order of the panels corresponds to the subsequent components of a term structure (maturities), whereas the yields corresponding to particular countries are illustrated by the row-wise order.

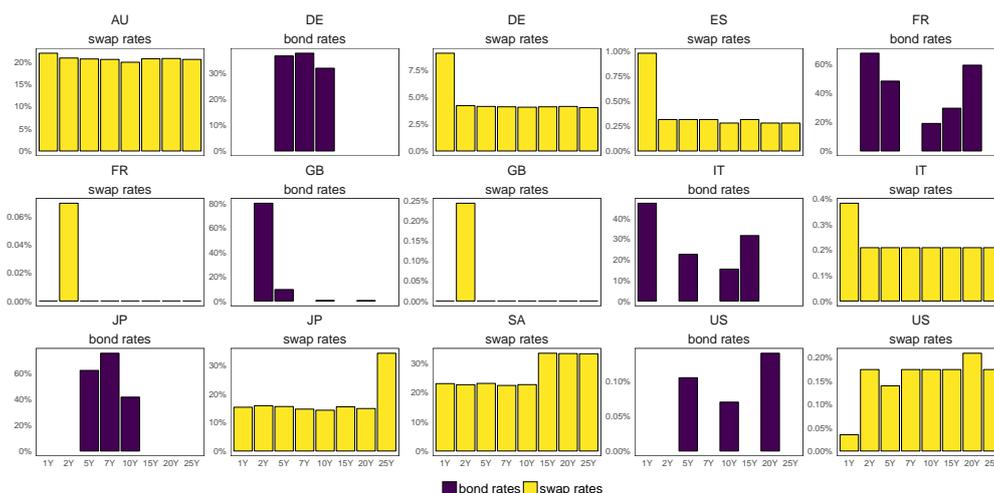


Figure S7. The proportion (%) of missing values present in the time series of inflation-linked sovereign bond or swap rates across different maturities and countries for a sample period 01/2006 - 12/2016 (2'870 days excluding weekends and holidays). The patterns of missigness are illustrated in Figure S8. The blank spaces on the x-axis correspond to instrumented which are not available in Bloomberg.

The reader may notice similarities between the country-specific yields from Figure S2 and the bond-based inflation-linked sovereign yield curves as the sovereign yield curves comprised either of conventional or inflation-linked bonds, share common dynamics which are characteristic among chosen countries and more resemble each other than the corresponding swap rates. On the other hand, the availability of the inflation-linked bond-based sovereign yield curves is much more limited both in time and the number of countries, than the swap-based yield curves. The panels in Figure S7 summarize the proportion of country-specific inflation-linked yield curves which are missing over the sample span, 01/2006-12/2016, which includes 2'870 days.

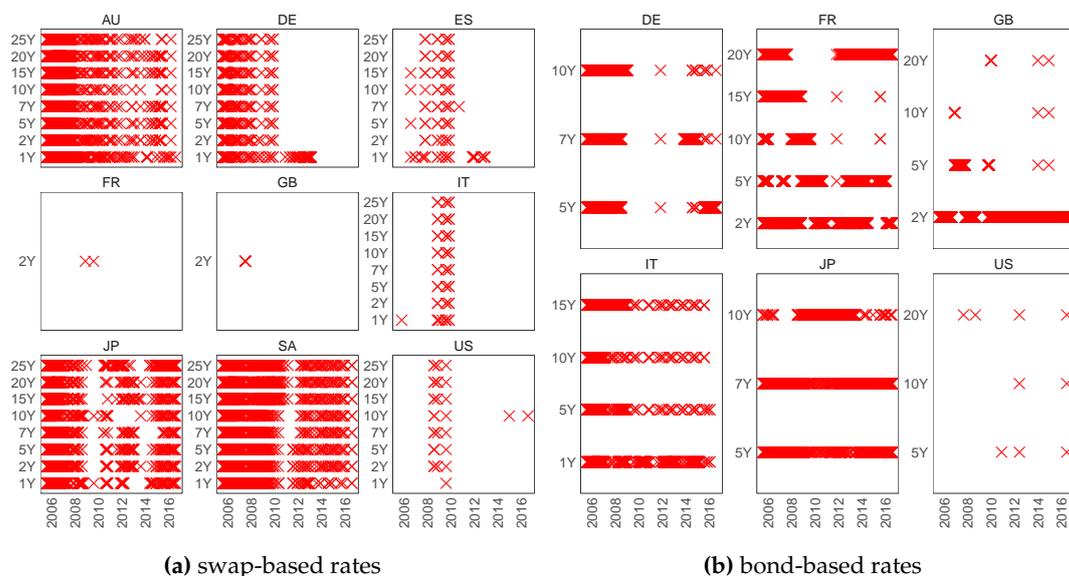


Figure S8. The indicators of missing values of swap-based (left) and sovereign (right) yield curves comprised of inflation-linked rates at different maturities for countries listed in Table 1. The red crosses on the left subfigure correspond to daily unavailabilities of rates. The missigness does not correspond to weekends and holidays.

A.4. Country-Specific Productivity Proxies

The country-specific productivity is a n i ndicator w hich m easures economic growth, competitiveness, and living standards within an economy. We consider productivity proxies from three indicators: yearly Gross Domestic Product in US dollars obtained from World Bank for 12 economies, quarterly time series of unemployment rates for 12 economies from Bloomberg and labour productivity for all countries except Brazil and South Africa from the OECD. In order to obtain labour productivity approximation for Brazil and South Africa, we collected quarterly numbers of GDP per labour force from OECD. The time series of GDP, Unemployment rates and labour productivity are presented in Figure S9 whereas the corresponding Bloomberg identifiers are listed in Table S3.

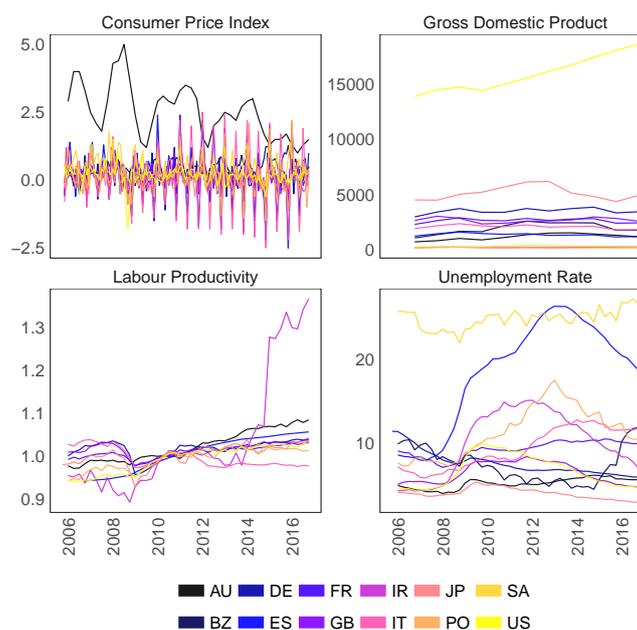


Figure S9. Daily Consumer Price Indexes (top left), quarterly Gross Domestic Products (top right), Unemployment rate in % (bottom left) and Labour Productivity (bottom right) of 12 countries listed in Table 1. Different colours of lines refer to different countries.

A.5. Euro-Zone Liquidity & Credit Proxies and Foreign Exchange

The instruments which we use as a proxy to Euro-Zone liquidity are Open Interests of Euro-Bund futures contracts and 2W Euro repo rate.

The Euro-Bund futures are the most popular contracts of EUREX, which is the largest exchange for Euro-denominated derivatives and has been ranked within main derivative exchanges by volume. The bond futures contracts are exchange-traded products with an underlying basket of deliverable bonds. The benchmark bonds of Euro-Bund futures are long-term notional debt securities issued by the German Federal Government at reference yield of 6% and with a term of 8.5 to 10.5 years which earliest two delivery months are available for trading. The bonds are considered the benchmark for long-term Euro-denominated government debts. The futures exist for 4 contracts which correspond to the following months: March, June, September and December. In the current study we use Generic 1st 'RX' Future contract provided by Bloomberg, that is, the time series of futures which is comprised of front-month instruments, the contracts with the shortest duration that could be purchased. As a proxy of Euro-zone liquidity, we use the open interest of Euro-Bund futures, the number of the outstanding contracts that exist on a given date. The amount of open interests represents the flow of the money to the futures market and indicates increased interests of market participants. The Figure S10a shows the dynamic of open interest over time (black solid line) with its smoothed curve. We used median

smoothing over a 20-day rolling window in order to remove spikes from the time series which are downward movements of a number of outstanding contracts when the instrument is close to its expiry date.

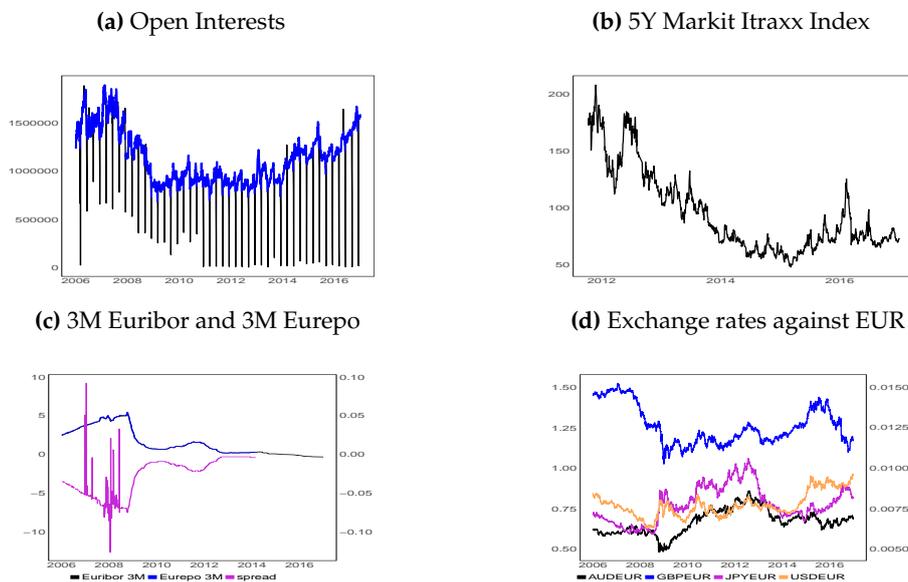


Figure S10. Top left panel (a): The raw (black) and smoothed (blue) open interests of Euro-Bund futures contracts over times. Top right panel (b): 5Y Marikt Itraxx Index over time. The bottom left panel (c): the spread (secondary y-axis) between 3 month Euribor and Eurorepo rates (%) over time. The bottom right panel (d): the exchange rates against EUR for AUD, GBP and USA (left y-axis) and JPY (right y-axis) over time.

In addition, as a proxy for short-term liquidity, we use the spread between the 3 months Euribor rate and the 3 months Eurorepo rate illustrated in Figure S10c, both reported by the European Banking Federation. The Euribor rate refers to Euro Interbank Offered Rate which determines, similarly as Euro Libor, the cost of short-term loans on the Euro-zone interbank market. The rate is published daily by European Bank Federation. The concept for the Euribor is similar to Euro Libor. Euribor is an average rate at which the biggest in volume European Union banks borrow funds from one another, whereas Euro Libor is determined by a group of banks which operate in London's money market. The 3 months Eurorepo rate refers to a benchmark repurchase rate which is offered by one bank to another bank for Euro-denominated deals. A repurchase agreement is a lending transaction with collateral, where one party agrees to sell securities to another party with the agreement to repurchase the same securities at a specific price in the future. The Bloomberg identifiers of rates are given in Table S3 in Appendix A.

The European Markit iTraxx indices trade 3, 5, 7 and 10-year maturities and are determined on the basis of liquidity every six months. The benchmark Markit iTraxx Europe index comprises 125 equally-weighted credit default swaps on investment grade European corporates. We chose as a credit proxy of Euro-zone the most actively traded contract at 5Y maturity. The index is presented in Figure S10d and its Bloomberg identifier is listed in Table S3.

The daily US Dollar (USD), Great Britain Pound (GBP) and Japanese Yen (JPN) exchange against Euro are plotted in Figure S10d. We obtained the rates by converting daily exchange rates AUDUSD, GBPUSD, USDJPY and EURUSD downloaded from Bloomberg to AUDEUR, GBPEUR, USDEUR and JPYEUR. The identifiers of instruments are listed in Table S3.

Appendix B. Supplementary Material to Section 6

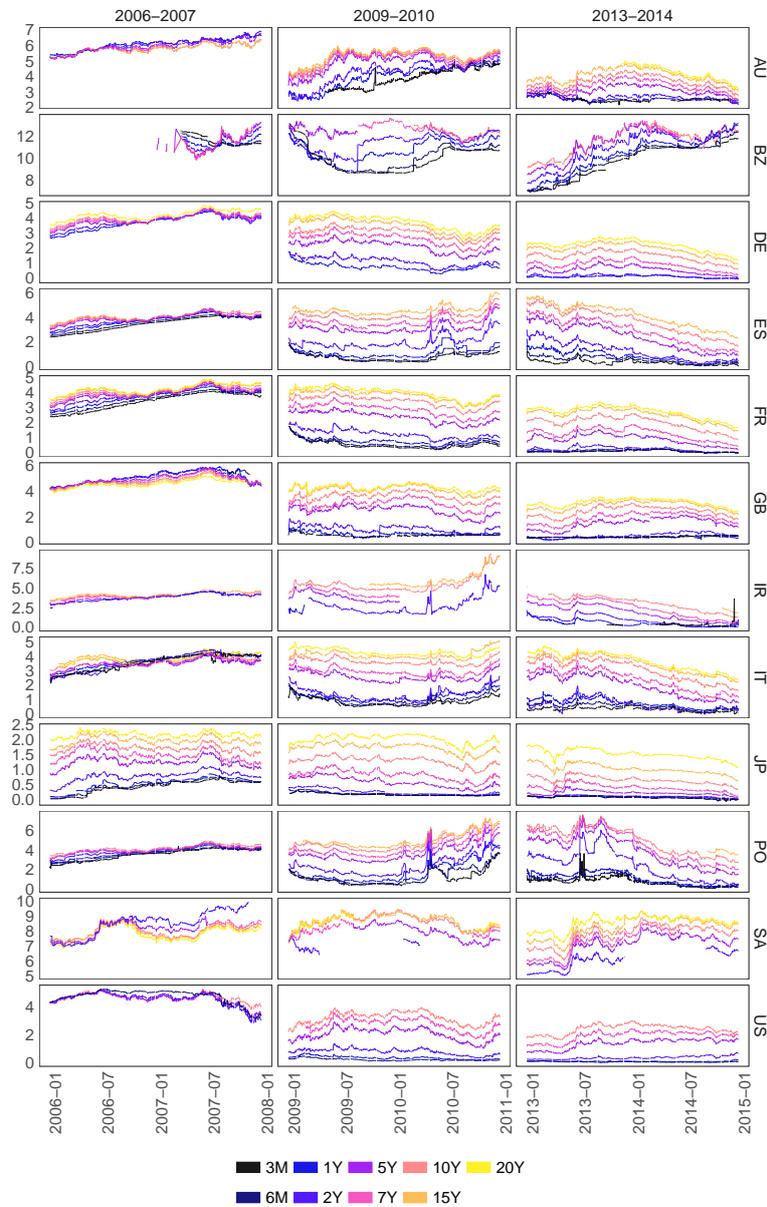


Figure S11. The yield curves of 12 countries (row-wise) from the data set **D2** for selected years (column-wise) across different maturities over time. The colors of lines correspond to the tenors and are listed in the legend.

Appendix C. **The EM Algorithm for Gaussian Probabilistic Principal Component Analysis and Its Robust Version**

The Gaussian PPCA assumes that X_t and τ_t are k - and d -dimensional zero mean Gaussian random vectors with covariance matrices equal to \mathbb{I}_k and $\sigma^2 \mathbb{I}_d$ respectively, which are mutually independent and independent over time. Given the model (1) of PPCA, Y_t is also d -dimensional random vector which follows Gaussian distribution and is independent over time, with mean μ and the covariance matrix $C = W^T W + \sigma^2 \mathbb{I}_d$. Please refer to Tipping and Bishop (1999) or Toczydlowska et al. (2017) for detailed derivations and discussion. The EM algorithm for Gaussian PPCA iterates over two steps, where the maximizers to the function Q of Gaussian PPCA are found in closed form. The Theorem 1 and Theorem 2 provide the formulations of E-step and M-step of EM algorithm respectively, assuming

the presence of missing data in realisation of the observation vector \mathbf{Y}_t which is partitioned into the subvectors corresponding to observed and missing entries, $\mathbf{Y}_t = [\mathbf{Y}_t^o, \mathbf{Y}_t^m]$. Recall, that the vector of static parameters of the algorithm is equal to $\Psi = [\boldsymbol{\mu}, \mathbf{W}, \sigma^2]$.

Theorem 1. The E-step of the EM algorithm for Gaussian PPCA given N realisations of the observation vector \mathbf{Y}_t^o denoted by $\mathbf{y}_{1:N}^o = \{\mathbf{y}_1^o, \dots, \mathbf{y}_N^o\}$ is given by

$$\begin{aligned} Q(\Psi, \Psi^*) &= \mathbb{E}_{\mathbf{Y}_{1:N}^m, \mathbf{X}_{1:N} | \mathbf{Y}_{1:N}^o, \Psi} \left[\log \left(\pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N} | \Psi^*}(\mathbf{y}_{1:N}, \mathbf{x}_{1:N}) \right) \right] \\ &= -\frac{1}{2} \sum_{t=1}^N \left\{ (d+k) \log 2\pi + d \log \sigma^{*2} + \frac{1}{\sigma^{*2}} \left(\text{tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t^T \mathbf{Y}_t \right] \right\} - 2 \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right] \boldsymbol{\mu}^{*T} \right. \right. \\ &\quad \left. \left. + \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T} - 2 \text{tr} \left\{ \mathbf{W}^* \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{X}_t^T \mathbf{Y}_t \right] \right\} + 2 \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{X}_t \right] \mathbf{W}^{*T} \boldsymbol{\mu}^{*T} \right) \right. \\ &\quad \left. + \text{tr} \left\{ \left(\frac{1}{\sigma^{*2}} \mathbf{W}^{*T} \mathbf{W}^* + \mathbb{I}_k \right) \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{X}_t^T \mathbf{X}_t \right] \right\} \right\} \end{aligned}$$

for the corresponding moments of the conditional distribution $\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi$

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right]_{1 \times d} &= \left[\boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) (\mathbf{W}_o \mathbf{W}_o^T + \sigma^2 \mathbb{I}_{d_o})^{-1} \mathbf{W}_o \mathbf{W}_m^T \right], \\ \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t^T \mathbf{Y}_t \right]_{d \times d} &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{mm} - \mathbf{W}_m \mathbf{W}_o^T \mathbf{C}_{oo}^{-1} \mathbf{W}_o \mathbf{W}_m^T \end{bmatrix} + \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right], \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{X}_t \right]_{1 \times k} &= \left(\mathbb{E}_{\mathbf{Y}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right] - \boldsymbol{\mu} \right) \mathbf{W} \mathbf{M}^{-1}, \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t | \mathbf{Y}_t^o, \Psi} \left[\mathbf{X}_t^T \mathbf{X}_t \right]_{k \times k} &= \sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \left(\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[(\mathbf{Y}_t - \boldsymbol{\mu})^T (\mathbf{Y}_t - \boldsymbol{\mu}) \right] \right) \mathbf{W} \mathbf{M}^{-1}, \\ \mathbb{E}_{\mathbf{X}_t, \mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{X}_t^T \mathbf{Y}_t \right]_{k \times d} &= \left(\mathbf{W}^T \mathbf{W} + \sigma^2 \mathbb{I}_k \right)^{-1} \mathbf{W}^T \left(\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t^T \mathbf{Y}_t \right] - \boldsymbol{\mu}^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right] \right). \end{aligned}$$

where $\mathbf{M} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbb{I}_k$.

Theorem 2. The maximizers of $Q(\Psi | \Psi^*)$ are the solution to the following set of the problems $\nabla_{\Psi^*} Q = \mathbf{0}$ and are given by

$$\begin{aligned} \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}} \left(\mathbb{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) + \boldsymbol{\mu} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\ \mathbf{W}^* &= \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*} \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\boldsymbol{\mu}} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \\ \sigma^{*2} &= \frac{1}{d} \text{tr} \left\{ \bar{\mathbf{C}}_{\boldsymbol{\mu}^*} - 2 \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*} + \mathbf{W}^{*T} \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\boldsymbol{\mu}} \mathbf{W} \mathbf{M}^{-1} \right) \right\} \end{aligned}$$

where

$$\begin{aligned} \bar{\boldsymbol{\mu}} &= \frac{1}{N} \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right], \\ \bar{\mathbf{S}} &= \frac{1}{N} \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} \left[\mathbf{Y}_t \right], \\ \bar{\mathbf{Q}} &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_m \mathbf{W}_m^T + \sigma^2 \mathbb{I}_{d_m} - \mathbf{W}_m \mathbf{W}_o^T \left(\mathbf{W}_o \mathbf{W}_o^T + \sigma^2 \mathbb{I}_{d_o} \right)^{-1} \mathbf{W}_o \mathbf{W}_m^T \end{bmatrix}, \\ \bar{\mathbf{C}}_{\boldsymbol{\mu}} &= \bar{\mathbf{S}} + \bar{\mathbf{Q}} - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}^{mT} \boldsymbol{\mu} + \boldsymbol{\mu}^T \boldsymbol{\mu}, \\ \bar{\mathbf{C}}_{\boldsymbol{\mu}^*} &= \bar{\mathbf{S}} + \bar{\mathbf{Q}} - \boldsymbol{\mu}^{*T} \bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}^{mT} \boldsymbol{\mu}^* + \boldsymbol{\mu}^{*T} \boldsymbol{\mu}^*, \end{aligned}$$

$$\bar{\mathbf{C}}_{\mu, \mu^*} = \bar{\mathbf{S}} + \bar{\mathbf{Q}} - \boldsymbol{\mu}^{*T} \bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}^m T \boldsymbol{\mu} + \boldsymbol{\mu}^{*T} \boldsymbol{\mu}.$$

The estimation of the parameters $\mathbf{W}, \boldsymbol{\mu}, \sigma^2$ follows the procedure in Algorithm 1. One can improved the robustness of Algorithm 1 by replacing sample mean $\bar{\boldsymbol{\mu}}$ and covariance $\bar{\mathbf{S}}$ by their robust estimates obtained, for instance, by employing S-estimators. The EM algorithm for Gaussian PPCA with the utilisation of robust estimators of location and scatter matrix is refereed as the EM algorithm for robust Gaussian PPCA.

Algorithm 1 EM Algorithm for Gaussian Probabilistic Principal Component Analysis

- 1: **for** $j = 1, \dots, d$ **do**
- 2: Compute $\Theta_N([\mathbf{Y}^o]_{\cdot,j}) = (\hat{\mu}_j, \hat{\sigma}_j^2)$
- 3: Standardize data $[\tilde{\mathbf{Y}}^o]_{\cdot,j} = \frac{[\mathbf{Y}^o]_{\cdot,j} - \hat{\mu}_j}{\hat{\sigma}_j}$
- 4: **end for**
- 5: $\tilde{\mathbf{Y}}^m = \mathbf{0}$ and $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}^o, \tilde{\mathbf{Y}}^m)$
- 6: Initialise: $\varepsilon, i = 0, \mathbf{W}^{(0)} = \mathbf{W}_0, \sigma^{2(0)} = \sigma_0^2,$
- 7: **repeat**
- 8: E-step:
- 9: **for** $t = 1, \dots, N$ **do**
- 10:
$$\mathbf{Y}_t^{(i)} = \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m^{(i)} + \left(\mathbf{Y}_t^o - \boldsymbol{\mu}_o^{(i)} \right) \left(\mathbf{W}_o^{(i)} \mathbf{W}_o^{(i)T} + \sigma^{2(i)} \mathbf{I}_{d_o} \right)^{-1} \mathbf{W}_o^{(i)} \mathbf{W}_m^{(i)T} \end{bmatrix}$$
- 11: **end for**
- 12: Calculate: $\bar{\boldsymbol{\mu}}^{(i)} = \frac{1}{N} \sum_{t=1}^N \mathbf{Y}_t^{(i)}, \bar{\mathbf{S}}^{(i)} = \frac{1}{N} \sum_{t=1}^N \mathbf{Y}_t^{(i)T} \mathbf{Y}_t^{(i)}$
- 13: M-step:

$$\begin{aligned} \boldsymbol{\mu}^{(i+1)} &= \bar{\boldsymbol{\mu}} \left(\mathbb{I}_d - \mathbf{W}^{(i)} \mathbf{M}^{(i)-1} \mathbf{W}^{(i)T} \right) + \boldsymbol{\mu}^{(i)} \mathbf{W}^{(i)} \mathbf{M}^{(i)T} \mathbf{W}^{(i+1)T} \\ \mathbf{W}^{(i+1)} &= \bar{\mathbf{C}}_{\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}^{(i+1)}} \mathbf{W}^{(i)} \mathbf{M}^{(i)-1} \left(\sigma^{2(i)} \mathbf{M}^{(i)-1} + \mathbf{M}^{(i)-1} \mathbf{W}^{(i)} \bar{\mathbf{C}}_{\boldsymbol{\mu}^{(i)}} \mathbf{W}^{(i)} \mathbf{M}^{(i)-1} \right)^{-1} \\ \sigma^{2(i+1)} &= \frac{1}{d} \text{Tr} \left\{ \bar{\mathbf{C}}_{\boldsymbol{\mu}^{(i+1)}} - 2 \mathbf{W}^{(i+1)} \mathbf{M}^{(i)-1} \mathbf{W}^{(i)T} \bar{\mathbf{C}}_{\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}^{(i+1)}} \right\} \\ &\quad + \frac{1}{d} \text{Tr} \left\{ \mathbf{W}^{(i+1)T} \mathbf{W}^{(i+1)} \left(\sigma^{2(i)} \mathbf{M}^{(i)-1} + \mathbf{M}^{(i)-1} \mathbf{W}^{(i)T} \bar{\mathbf{C}}_{\boldsymbol{\mu}^{(i)}} \mathbf{W}^{(i)} \mathbf{M}^{(i)-1} \right) \right\} \end{aligned}$$

- 14: $i = i + 1$
 - 15: **until** a convergence criterion is satisfied
-

Appendix D. Proofs of The Theorems in Section 4

In the following section we provide the proofs to subsequent steps of EM algorithms for two cases of t-Student PPCA discussed in Section 4. First subsection lists the essential notation used in the proofs. Next we shows the steps of the proofs to the two theorems from Section 4.1 and later to the theorems in Section 4.2, for t-Student IND and t-Student IID PPCA frameworks, respectively.

D.1. Notation

Let us recall the general notation. The bold capital letters denote matrices, the bold small letters refer to vectors and small letters to scalars. The random variables are denoted by capital letters usually with a lower index corresponding to point in time such as U_t . In addition, the random vectors are bold such i.e. \mathbf{Y}_t . If a capital letter without the lower index is used to denote random variable, it is highlighted in the text i.e. U . The realisations of the random variables are denoted by the small letter such as \mathbf{y}_t being a realisation of \mathbf{Y}_t and u of U .

We introduce the following notation which is used in the proofs of the theorems stated in the study.

$$= \sum_{t=1}^N \int_{\mathbb{R}_+} \int_{\mathbb{R}^k} \int_{\mathbb{R}^{dm}} \log \left(\pi_{\mathbf{Y}_t, \mathbf{X}_t, U_t | \Psi^*}(\mathbf{y}_t, \mathbf{x}_t, u_t) \right) \pi_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi}(\mathbf{y}_t^m, \mathbf{x}_t, u_t) d\mathbf{y}_t^m d\mathbf{x}_t du_t$$

Hence we have that

$$\begin{aligned} Q(\Psi, \Psi^*) &= \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} \left[\log \pi_{\mathbf{Y}_t, \mathbf{X}_t, U_t | \Psi^*}(\mathbf{y}_t, \mathbf{x}_t, u_t) \right] \\ &= -\frac{1}{2} \sum_{t=1}^N \left\{ \left(d + k - \frac{v^*}{2} \right) \log 2\pi + d \log \sigma^{*2} - (d + k + v^* - 2) \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [\log U_t] \right. \\ &\quad \left. + \frac{1}{\sigma^{*2}} \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*) (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T] - \frac{2}{\sigma^{*2}} \text{tr} \left\{ \mathbf{W}^* \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} \right. \\ &\quad \left. + \text{tr} \left\{ \left(\frac{1}{\sigma^{*2}} \mathbf{W}^{*T} \mathbf{W}^* + \mathbb{I}_k \right) \mathbb{E}_{\mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{X}_t] \right\} + v^* \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] - 2 \log \left(\frac{\left(\frac{v^*}{2} \right)^{\frac{v^*}{2}}}{\Gamma \left(\frac{v^*}{2} \right)} \right) \right\} \\ &= -\frac{1}{2} \sum_{t=1}^N \left\{ \left(d + k - \frac{v^*}{2} \right) \log 2\pi + d \log \sigma^{*2} - (d + k + v^* - 2) \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [\log U_t] \right. \\ &\quad \left. + \frac{1}{\sigma^{*2}} \text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} \right. \\ &\quad \left. - \frac{2}{\sigma^{*2}} \text{tr} \left\{ \mathbf{W}^* \left(\mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{Y}_t] - \mathbb{E}_{\mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T] \boldsymbol{\mu}^* \right) \right\} \right. \\ &\quad \left. + \text{tr} \left\{ \left(\frac{1}{\sigma^{*2}} \mathbf{W}^{*T} \mathbf{W}^* + \mathbb{I}_k \right) \mathbb{E}_{\mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{X}_t] \right\} + v^* \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] - 2 \log \left(\frac{\left(\frac{v^*}{2} \right)^{\frac{v^*}{2}}}{\Gamma \left(\frac{v^*}{2} \right)} \right) \right\}. \end{aligned}$$

Since the distribution $\mathbf{Y}_t | U_t, \Psi$ is Gaussian, it can be easily shown that the distribution of the vector \mathbf{Y}_t conditioned on the subvector of the observed values and the variable U_t , is specified as follows

$$\mathbf{Y}_t | \mathbf{Y}_t^o, U_t \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix}, \frac{1}{U_t} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{mm} - \mathbf{C}_{mo} \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix} \right)$$

where $\mathbf{C}_{oo} = \mathbf{W}_o \mathbf{W}_o^T + \sigma^2 \mathbb{I}_{d_o}$ and $\mathbf{C}_{mm} = \mathbf{W}_m \mathbf{W}_m^T + \sigma^2 \mathbb{I}_{d_m}$ are square, symmetric submatrices of $\mathbf{C} = \mathbf{W} \mathbf{W}^T + \sigma^2 \mathbb{I}_d$ with rows (left side index) and columns (right side index) corresponding to observed entries and missing entries, respectively. The matrices $\mathbf{C}_{om} = \mathbf{W}_o \mathbf{W}_m^T$ is a submatrix of the matrix \mathbf{C} which rows correspond to observed entries and columns to missing entries of the vector \mathbf{Y}_t , and $\mathbf{C}_{mo} = \mathbf{C}_{om}^T$. Let us denote

$$\mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{mm} - \mathbf{C}_{mo} \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix}$$

We can specify the following moments which are useful to derive the conditional moments of the distribution $\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o$

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t]_{1 \times d} &= \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix}, \\ \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t^T \mathbf{Y}_t]_{d \times d} &= \frac{1}{U_t} \mathbf{Q} + \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, \Psi} [\mathbf{Y}_t]. \end{aligned}$$

Let us define $D_t^o = (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{C}_{oo}^{-1} (\mathbf{Y}_t^o - \boldsymbol{\mu}_o)^T$ being a squared Mahalanobias distance of the observed data subvector \mathbf{Y}_t^o with marginal distribution $\mathbf{Y}_t^o \sim \mathcal{N}(\boldsymbol{\mu}_o, \mathbf{C}_{oo})$. The moments of the variables $U_t | \mathbf{Y}_t^o, \Psi$

and $\log U_t | \mathbf{Y}_t^o, \Psi$ are calculated similarly as their equivalents in complete data setting, using the marginal distribution of \mathbf{Y}_t^o , and are given by

$$\begin{aligned} \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] &= \frac{v + d_o}{v + D_t^o}, \\ \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [\log U_t] &= \psi\left(\frac{v + d_o}{2}\right) - \log\left(\frac{v + D_t^o}{2}\right). \end{aligned}$$

The moments of the joint distribution of the latent variables $\mathbf{X}_t, \mathbf{Y}_t^m$ and U_t conditioned on the subvector of observed entries \mathbf{Y}_t^o are obtained using the Law of Total Expectation. Applying the chain rule of probability to the joint distribution

$$\pi_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} = \pi_{\mathbf{X}_t | \mathbf{Y}_t, U_t, \Psi} \cdot \pi_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} \cdot \pi_{U_t | \mathbf{Y}_t^o, \Psi}$$

and the fact that conditional means of $\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t$ and $\mathbf{X}_t | \mathbf{Y}_t^o, U_t$ are not dependent on U_t , we obtain

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t]_{1 \times d} &= \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t] = \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} \left[U_t \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t] \right] = \\ &= \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix} \\ &= \frac{v + d_o}{v + D_t^o} \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix}, \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t^T \mathbf{Y}_t]_{d \times d} &= \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t^T \mathbf{Y}_t] = \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} \left[U_t \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t^T \mathbf{Y}_t] \right] = \\ &= \mathbf{Q} + \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t] \\ &= \mathbf{Q} + \frac{v + d_o}{v + D_t^o} \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t], \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t]_{1 \times k} &= \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} \left[U_t \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbb{E}_{\mathbf{X}_t | \mathbf{Y}_t, U_t, \Psi} (\mathbf{X}_t)] \right] \\ &= \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t]] \mathbf{W} \mathbf{M}^{-1} = \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t] \mathbf{W} \mathbf{M}^{-1} \\ &= \frac{v + d_o}{v + D_t^o} \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{C}_{oo}^{-1} \mathbf{C}_{om} \end{bmatrix} \mathbf{W} \mathbf{M}^{-1}, \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{X}_t]_{k \times k} &= \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} \left[U_t \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbb{E}_{\mathbf{X}_t | \mathbf{Y}_t, U_t, \Psi} [\mathbf{X}_t^T \mathbf{X}_t]] \right] \\ &= \sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu})^T (\mathbf{Y}_t - \boldsymbol{\mu})] \mathbf{W} \mathbf{M}^{-1}, \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu})^T (\mathbf{Y}_t - \boldsymbol{\mu})]_{d \times d} &= \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu})^T (\mathbf{Y}_t - \boldsymbol{\mu})] \\ &= \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t^T \mathbf{Y}_t] - 2 \boldsymbol{\mu}^T \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t] + \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] \boldsymbol{\mu}^T \boldsymbol{\mu}, \\ \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t^T \mathbf{X}_t]_{d \times k} &= \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} \left[U_t \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [\mathbf{Y}_t^T \mathbb{E}_{\mathbf{X}_t | \mathbf{Y}_t, U_t, \Psi} [\mathbf{X}_t]] \right] \\ &= \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} \left[\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U_t, \Psi} [U_t \mathbf{Y}_t^T (\mathbf{Y}_t - \boldsymbol{\mu})] \right] \mathbf{W} \mathbf{M}^{-1} \\ &= \left(\mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t^T \mathbf{Y}_t] - \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{Y}_t]^T \boldsymbol{\mu} \right) \mathbf{W} \mathbf{M}^{-1}. \end{aligned}$$

□

D.2.2. The Proof of Theorem 2

Proof. Using the notation introduced in Theorem 2, the maximizers of the function

$$\begin{aligned}
 Q(\Psi, |\Psi^*) &= -\frac{1}{2} \sum_{t=1}^N \left\{ \left(d + k - \frac{v^*}{2} \right) \log 2\pi + d \log \sigma^{*2} - (d + k + v^* - 2) \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [\log U_t] \right. \\
 &\quad \left. + \frac{1}{\sigma^{*2}} \text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} \right. \\
 &\quad \left. - \frac{2}{\sigma^{*2}} \text{tr} \left\{ \mathbf{W}^* \left(\mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{Y}_t] - \mathbb{E}_{\mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t] \right)^T \boldsymbol{\mu}^* \right\} \right. \\
 &\quad \left. + \text{tr} \left\{ \left(\frac{1}{\sigma^{*2}} \mathbf{W}^{*T} \mathbf{W}^* + \mathbb{I}_k \right) \mathbb{E}_{\mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{X}_t] \right\} + v^* \mathbb{E}_{U_t | \mathbf{Y}_t^o, \Psi} [U_t] - 2 \log \left(\frac{\left(\frac{v^*}{2} \right)^{\frac{v^*}{2}}}{\Gamma \left(\frac{v^*}{2} \right)} \right) \right\}
 \end{aligned}$$

with respect to the vector of the static parameters $(\mathbf{W}^*, \boldsymbol{\mu}^*, \sigma^{*2})$ which are the solutions to the set of equations $\nabla_{\Psi^*} Q = \mathbf{0}$ and are given by

$$\begin{aligned}
 \frac{\partial Q(\Psi, \Psi^*)}{\partial \boldsymbol{\mu}^*} &= -\frac{1}{2\sigma^{*2}} \sum_{t=1}^N \left\{ -2 \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)] + 2 \mathbb{E}_{\mathbf{X}_t, \mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t] \mathbf{W}^{*T} \right\} = \mathbf{0} \\
 \iff \sum_{t=1}^N \left\{ \frac{v + d_o}{v + D_t^o} \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [\mathbf{Y}_t] - \boldsymbol{\mu}^* \right) - \frac{v + d_o}{v + D_t^o} \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [\mathbf{Y}_t] - \boldsymbol{\mu} \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right\} &= \mathbf{0} \\
 \iff \underbrace{\frac{1}{N} \sum_{t=1}^N \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [\mathbf{Y}_t] \frac{v + d_o}{v + D_t^o} \right)}_{\bar{\boldsymbol{\mu}}_{ts}} \left(\mathbb{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) + \underbrace{\frac{1}{N} \sum_{t=1}^N \left(\frac{v + d}{v + D_t^o} \right)}_{\bar{u}} \boldsymbol{\mu} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\
 - \boldsymbol{\mu}^* \frac{1}{N} \sum_{t=1}^N \frac{v + d}{v + D_t^o} &= \mathbf{0} \\
 \iff \boldsymbol{\mu}^* &= \frac{1}{\bar{u}(\mathbf{y}_{1:N}^o; \Psi)} \left(\bar{\boldsymbol{\mu}}_{ts}(\mathbf{y}_{1:N}^o; \Psi) \left(\mathbb{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) + \bar{u}(\mathbf{y}_{1:N}^o; \Psi) \boldsymbol{\mu} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial Q(\Psi, \Psi^*)}{\partial \mathbf{W}^*} &= -\frac{1}{2\sigma^{*2}} \sum_{t=1}^N \left\{ -2 \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T \mathbf{X}_t] \right. \\
 &\quad \left. + 2 \mathbf{W}^* \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^o, \Psi} [U_t \mathbf{X}_t^T \mathbf{X}_t] \right\} = \mathbf{0} \\
 \iff \sum_{t=1}^N \left\{ \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu})] \mathbf{W} \mathbf{M}^{-1} \right. \\
 &\quad \left. - \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu})^T (\mathbf{Y}_t - \boldsymbol{\mu})] \mathbf{W} \mathbf{M}^{-1} \right) \right\} = \mathbf{0} \\
 \iff \frac{1}{N} \sum_{t=1}^N \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu})] \right) \mathbf{W} \mathbf{M}^{-1} \\
 \underbrace{\hspace{10em}}_{\mathbf{C}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}^{ts}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \\
 - \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \frac{1}{N} \sum_{t=1}^N \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^o, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu})^T (\mathbf{Y}_t - \boldsymbol{\mu})] \right) \mathbf{W} \mathbf{M}^{-1} \right) &= \mathbf{0} \\
 \underbrace{\hspace{10em}}_{\mathbf{C}_{\boldsymbol{\mu}}^{ts}(\mathbf{y}_{1:N}^o; \Psi)}
 \end{aligned}$$

$$\begin{aligned} &\iff \mathbf{W}^* = \bar{\mathbf{C}}_{\mu, \mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts}(\mathbf{y}_{1:N}^0; \Psi) \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \\ \frac{\partial Q(\Psi, \Psi^*)}{\partial \sigma^{*2}} &= -\frac{1}{2} \sum_{t=1}^N \left\{ \frac{d}{\sigma^{*2}} - \frac{1}{(\sigma^{*2})^2} \left(\text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} \right. \right. \\ &\quad \left. \left. - 2 \text{Tr} \left\{ \mathbf{W}^* \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^0, \Psi} [U_t \mathbf{X}_t^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} + \text{Tr} \left\{ \mathbf{W}^{*T} \mathbf{W}^* \mathbb{E}_{\mathbf{Y}_t^m, \mathbf{X}_t, U_t | \mathbf{Y}_t^0, \Psi} [U_t \mathbf{X}_t^T \mathbf{X}_t] \right\} \right) \right\} = 0 \\ &\iff \sum_{t=1}^N \left\{ \frac{d}{\sigma^{*2}} - \frac{1}{(\sigma^{*2})^2} \left(\text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} \right. \right. \\ &\quad \left. \left. - 2 \text{Tr} \left\{ \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right\} \right. \right. \\ &\quad \left. \left. + \text{Tr} \left\{ \mathbf{W}^{*T} \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \mathbf{W} \mathbf{M}^{-1} \right) \right\} \right) \right\} = 0 \\ &\iff \frac{d}{\sigma^{*2}} - \frac{1}{(\sigma^{*2})^2} \text{Tr} \left\{ \underbrace{\frac{1}{N} \sum_{t=1}^N \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right)}_{\bar{\mathbf{C}}_{\mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*)} \right. \\ &\quad \left. - 2 \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \underbrace{\frac{1}{N} \sum_{t=1}^N \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right)}_{\bar{\mathbf{C}}_{\mu, \mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*)} \right. \\ &\quad \left. + \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \underbrace{\frac{1}{N} \sum_{t=1}^N \left(\mathbb{E}_{\mathbf{Y}_t^m, U_t | \mathbf{Y}_t^0, \Psi} [U_t (\mathbf{Y}_t - \boldsymbol{\mu}^*)^T (\mathbf{Y}_t - \boldsymbol{\mu}^*)] \right)}_{\bar{\mathbf{C}}_{\mu}^{ts}(\mathbf{y}_{1:N}^0; \Psi)} \right) \mathbf{W} \mathbf{M}^{-1} \right) \mathbf{W}^{*T} \right\} = 0 \\ &\iff \sigma^{*2} = \frac{1}{d} \text{Tr} \left\{ \bar{\mathbf{C}}_{\mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*) - 2 \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu, \mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*) \right\} \\ &\quad + \text{Tr} \left\{ \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts}(\mathbf{y}_{1:N}^0; \Psi) \mathbf{W} \mathbf{M}^{-1} \right) \mathbf{W}^{*T} \right\} \end{aligned}$$

where

$$\begin{aligned} \bar{u}(\mathbf{y}_{1:N}^0; \Psi) &= \frac{1}{N} \sum_{t=1}^N \frac{v + d_o}{v + D_t^o} \\ \bar{\boldsymbol{\mu}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) &= \frac{1}{N} \sum_{t=1}^N \frac{v + d_o}{v + D_t^o} \mathbb{E}_{\mathbf{Y}_t | \mathbf{Y}_t^0, U_t, \Psi} [\mathbf{Y}_t], \\ \bar{\mathbf{S}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) &= \frac{1}{N} \sum_{t=1}^N \frac{v + d_o}{v + D_t^o} \mathbb{E}_{\mathbf{Y}_t | \mathbf{Y}_t^0, U_t, \Psi} [\mathbf{Y}_t] \mathbb{E}_{\mathbf{Y}_t | \mathbf{Y}_t^0, U_t, \Psi} [\mathbf{Y}_t]^T, \\ \bar{\mathbf{C}}_{\mu}^{ts}(\mathbf{y}_{1:N}^0; \Psi) &= \bar{\mathbf{S}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) + \mathbf{Q} - 2\boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) + \bar{u}(\mathbf{y}_{1:N}^0; \Psi) \boldsymbol{\mu}^T \boldsymbol{\mu}, \\ \bar{\mathbf{C}}_{\mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*) &= \bar{\mathbf{S}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) + \mathbf{Q} - 2\boldsymbol{\mu}^{*T} \bar{\boldsymbol{\mu}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) + \bar{u}(\mathbf{y}_{1:N}^0; \Psi) \boldsymbol{\mu}^{*T} \boldsymbol{\mu}^*, \\ \bar{\mathbf{C}}_{\mu, \mu^*}^{ts}(\mathbf{y}_{1:N}^0; \Psi, \Psi^*) &= \bar{\mathbf{S}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) + \mathbf{Q} - (\boldsymbol{\mu}^* + \boldsymbol{\mu})^T \bar{\boldsymbol{\mu}}_{ts}(\mathbf{y}_{1:N}^0; \Psi) + \bar{u}(\mathbf{y}_{1:N}^0; \Psi) \boldsymbol{\mu}^{*T} \boldsymbol{\mu}. \end{aligned}$$

The reader may notice that the vector $\boldsymbol{\mu}^*$ is a linear function \mathbf{W}^* and vice versa. Hence, we can solve the following linear system of equations

$$\boldsymbol{\mu}^* = \frac{1}{\bar{u}} \left(\bar{\boldsymbol{\mu}}_{ts} \left(\mathbb{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) + \bar{u} \boldsymbol{\mu} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right)$$

$$\mathbf{W}^* = \bar{\mathbf{C}}_{\mu, \mu^*}^{ts} \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1}$$

by substituting the values of \mathbf{W}^* in the top equation by its values in the bottom equation, that is

$$\begin{aligned} \bar{\boldsymbol{\mu}} \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}_{ts} + (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \underbrace{\left(\bar{\mathbf{C}}_{\mu, \mu^*}^{ts} \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \right)^T}_{\mathbf{W}^*} \\ \bar{\boldsymbol{\mu}} \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}_{ts} + (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu, \mu^*}^{ts}. \end{aligned}$$

Next we substitute $\bar{\mathbf{C}}_{\mu, \mu^*}^{ts} = \bar{\mathbf{S}}_{ts} + \mathbf{Q} - \boldsymbol{\mu}^{*T} \bar{\boldsymbol{\mu}}_{ts} - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}_{ts}^m + \bar{\boldsymbol{\mu}} \boldsymbol{\mu}^{*T}$ which contains the vector $\boldsymbol{\mu}^*$. We obtain that

$$\begin{aligned} \bar{\boldsymbol{\mu}} \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}_{ts} + (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\mathbf{S}}_{ts} + \mathbf{Q} - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}_{ts} \right) \\ &\quad + (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\boldsymbol{\mu}} \boldsymbol{\mu}^{*T} - \boldsymbol{\mu}^{*T} \bar{\boldsymbol{\mu}}_{ts} \right) \\ &\iff \\ &\left(\bar{\boldsymbol{\mu}} - (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts} \right)^T \right) \boldsymbol{\mu}^* \\ &= \bar{\boldsymbol{\mu}}_{ts} + (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\mathbf{S}}_{ts} + \mathbf{Q} - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}_{ts} \right) \\ &\iff \\ \boldsymbol{\mu}^* &= \left(\bar{\boldsymbol{\mu}} - (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts} \right)^T \right)^{-1} \\ &\quad \times \left(\bar{\boldsymbol{\mu}}_{ts} + (\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}_{ts}) \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\mathbf{S}}_{ts} + \mathbf{Q} - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}_{ts} \right) \right) \end{aligned}$$

for

$$\begin{aligned} \mathbf{W}^* &= \bar{\mathbf{C}}_{\mu, \mu^*}^{ts} \mathbf{W} \mathbf{M}^{-1} \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right)^{-1} \\ \sigma^{*2} &= \frac{1}{d} \text{Tr} \left\{ \bar{\mathbf{C}}_{\mu^*}^{ts} - 2 \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu, \mu^*}^{ts} + \mathbf{W}^* \left(\sigma^2 \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{C}}_{\mu}^{ts} \mathbf{W} \mathbf{M}^{-1} \right) \mathbf{W}^{*T} \right\} \end{aligned}$$

□

D.3. Identical and Conditionally Independent t-Student Case

In the following subsection we first provide a range of proposition in Subsubsection D.3.1 which are used as a stepping stones to the larger results, that is the proof of theorems stated in Section 4.2, to Theorem 3 in Appendix D.3.2 and to Theorem 4 in Appendix D.3.3.

D.3.1. Supporting Lemmas

Lemma 1. Let a d dimensional observation vector \mathbf{Y}_t be modelled as in Equation (1) with the assumptions given in Equation (4). The function $h : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}$ of single realisations of \mathbf{Y}_t and a Gamma random variable U

under the assumption of the Identical and Conditionally Independent t-Student case from Section 3.2 is defined as

$$h(\mathbf{y}_t, u; \Psi) = \pi_{\mathbf{Y}_t|\mathbf{X}_t, U, \Psi}(\mathbf{y}_t) \cdot \pi_{\mathbf{X}_t|U, \Psi}(\mathbf{x}_t) \cdot \varphi(\mathbf{x}; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}})^{-1} \\ = (2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d-k}{2}} u^{\frac{d}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{u}{2\sigma^2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\}$$

for $\tilde{\boldsymbol{\mu}}_t, 1 \times k = (\mathbf{y}_t - \boldsymbol{\mu})\mathbf{W}\mathbf{M}^{-1}$, $\tilde{\boldsymbol{\Sigma}}_{k \times k} = \frac{\sigma^2}{u}\mathbf{M}^{-1}$, $\mathbf{M}_{k \times k} = \mathbf{W}^T\mathbf{W} + \sigma^2\mathbb{I}_k$ and $\varphi(\mathbf{x}; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}})$ being a density function of k dimensional Gaussian random vector \mathbf{X} with a mean vector $\tilde{\boldsymbol{\mu}}_t$ and a covariance matrix $\tilde{\boldsymbol{\Sigma}}$.

Proof. Given the assumptions of the identically and conditionally independently distributed \mathbf{X}_t and ϵ and their stochastic representations from Section 3.2, $\pi_{\mathbf{Y}_t|\mathbf{X}_t, U, \Psi} \cdot \pi_{\mathbf{X}_t|U, \Psi}$ is a convolution of two Gaussian densities, that is

$$\pi_{\mathbf{Y}_t|\mathbf{X}_t, U, \Psi}(\mathbf{y}_t) \cdot \pi_{\mathbf{X}_t|U, \Psi}(\mathbf{x}_t) = (2\pi)^{-\frac{k+d}{2}} (\sigma^2)^{-\frac{d}{2}} u^{\frac{d+k}{2}} \exp \left\{ -\frac{u}{2\sigma^2} \left((\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{x}_t\mathbf{W}^T) (\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{x}_t\mathbf{W}^T)^T + \sigma^2\mathbf{x}_t\mathbf{x}_t^T \right) \right\} \\ = (2\pi)^{-\frac{k+d}{2}} (\sigma^2)^{-\frac{d}{2}} u^{\frac{d+k}{2}} \exp \left\{ -\frac{u}{2\sigma^2} \left(\mathbf{x}_t (\mathbf{W}^T\mathbf{W} + \sigma^2\mathbb{I}_k) \mathbf{x}_t^T - 2\mathbf{x}_t\mathbf{W}^T (\mathbf{y}_t - \boldsymbol{\mu})^T + (\mathbf{y}_t - \boldsymbol{\mu}) (\mathbf{y}_t - \boldsymbol{\mu})^T \right) \right\} \\ = (2\pi)^{-\frac{k+d}{2}} (\sigma^2)^{-\frac{d}{2}} u^{\frac{d+k}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_t) \tilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_t)^T \right\} \exp \left\{ -\frac{1}{2} \left(\frac{u}{\sigma^2} (\mathbf{y}_t - \boldsymbol{\mu}) (\mathbf{y}_t - \boldsymbol{\mu})^T - \tilde{\boldsymbol{\mu}}_t \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{\mu}}_t^T \right) \right\}$$

where $\tilde{\boldsymbol{\mu}}_t, 1 \times k = (\mathbf{y}_t - \boldsymbol{\mu})\mathbf{W}\mathbf{M}^{-1}$, $\tilde{\boldsymbol{\Sigma}}_{k \times k} = \frac{\sigma^2}{u}\mathbf{M}^{-1}$ and $\mathbf{M}_{k \times k} = \mathbf{W}^T\mathbf{W} + \sigma^2\mathbb{I}_k$. Let us define $\varphi(\mathbf{x}; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}})$ as a density function of a Gaussian random variable $\mathbf{X} \in \mathbb{R}^k$ with a vector of means $\tilde{\boldsymbol{\mu}}_t$ and a covariance matrix $\tilde{\boldsymbol{\Sigma}}$. We have that

$$\pi_{\mathbf{Y}_t|\mathbf{X}_t, U, \Psi}(\mathbf{y}_t) \cdot \pi_{\mathbf{X}_t|U, \Psi}(\mathbf{x}_t) = (2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d}{2}} u^{\frac{d+k}{2}} |\tilde{\boldsymbol{\Sigma}}|^{\frac{1}{2}} \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) \times \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\frac{u}{\sigma^2} \mathbb{I}_d - \mathbf{W}\tilde{\boldsymbol{\Sigma}}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\} \\ = (2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d-k}{2}} u^{\frac{d}{2}} |\mathbf{M}|^{-\frac{1}{2}} \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\frac{u}{\sigma^2} \mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\} \\ = (2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d-k}{2}} u^{\frac{d}{2}} |\mathbf{M}|^{-\frac{1}{2}} \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) \exp \left\{ -\frac{u}{2\sigma^2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\} \\ = h(\mathbf{y}_t, u; \Psi) \cdot \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}})$$

where function $h : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined as

$$h(\mathbf{y}_t, u; \Psi) = (2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d-k}{2}} u^{\frac{d}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{u}{2\sigma^2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\}.$$

□

Lemma 2. Let a d dimensional observation vector \mathbf{Y}_t be modelled as in Equation (1) with the assumptions given in Equation (4). The function $w : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}$ of single realisations of \mathbf{Y}_t and a Gamma random variable U under the assumption of the Identical and Conditionally Independent t-Student case from Section 3.2 is defined as

$$w(\mathbf{y}_t, u; \Psi, \Psi^*) = \int_{\mathbb{R}^k} \log \left(\pi_{\mathbf{Y}_t, \mathbf{X}_t|U, \Psi^*}(\mathbf{y}_t, \mathbf{x}_t) \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t = -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u$$

$$\begin{aligned}
 & - \frac{u}{2\sigma^{*2}} \text{Tr} \left\{ (\mathbf{y}_t - \boldsymbol{\mu}^*)^T (\mathbf{y}_t - \boldsymbol{\mu}^*) - 2(\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \\
 & - \frac{1}{2\sigma^{*2}} \text{Tr} \left\{ (u \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}) \mathbf{W} \mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1}) (\mathbf{W}^{*T} \mathbf{W}^* + \sigma^{*2} \mathbb{I}_k) \right\}
 \end{aligned}$$

for $\tilde{\boldsymbol{\mu}}_t, 1 \times k = (\mathbf{y}_t - \boldsymbol{\mu}) \mathbf{W} \mathbf{M}^{-1}$, $\tilde{\boldsymbol{\Sigma}}_{k \times k} = \frac{\sigma^2}{u} \mathbf{M}^{-1}$, $\mathbf{M}_{k \times k} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbb{I}_k$ and $\varphi(\mathbf{x}; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}})$ being a distribution function of k dimensional Gaussian random vector \mathbf{X} with mean $\tilde{\boldsymbol{\mu}}_t$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$.

Proof. Given the definition of \mathbf{X}_t and ϵ_t from Equation (4) and the complete data likelihood from Equation (5) of the Identical and Conditionally Independent t-Student case, let us define function $w : \mathbb{R}^d \times \mathbb{R}_+ \leftarrow \mathbb{R}$ as a solution to the following integration problem

$$w(\mathbf{y}_t, u; \Psi, \Psi^*) = \int_{\mathbb{R}^k} \log \left(\pi_{\mathbf{Y}_t, \mathbf{X}_t, U | \Psi^*}(\mathbf{y}_t, \mathbf{x}_t, u) \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t$$

where $\tilde{\boldsymbol{\mu}}_t, 1 \times k = (\mathbf{y}_t - \boldsymbol{\mu}) \mathbf{W} \mathbf{M}^{-1}$, $\tilde{\boldsymbol{\Sigma}}_{k \times k} = \frac{\sigma^2}{u} \mathbf{M}^{-1}$ and $\mathbf{M}_{k \times k} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbb{I}_k$ are specified as in Lemma 1. The function $\varphi(\mathbf{x}; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}})$ is a density function of the k -dimensional Gaussian random variable $\mathbf{X} \in \mathbb{R}^k$ with a vector of means $\tilde{\boldsymbol{\mu}}_t$ and a covariance matrix $\tilde{\boldsymbol{\Sigma}}$. Therefore, using the Chain Rule of Probability and the fact that \mathbf{Y}_t and \mathbf{X}_t are conditionally Gaussian, we calculate the function w as follows

$$\begin{aligned}
 w(\mathbf{y}_t, u; \Psi, \Psi^*) &= \int_{\mathbb{R}^k} \log \left(\pi_{\mathbf{Y}_t | \mathbf{X}_t, U, \Psi^*}(\mathbf{y}_t) \cdot \pi_{\mathbf{X}_t | U, \Psi^*}(\mathbf{x}_t) \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t \\
 &= \int_{\mathbb{R}^k} \log \left((2\pi)^{-\frac{d+k}{2}} (\sigma^{*2})^{-\frac{d}{2}} u^{\frac{d+k}{2}} \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t \\
 &\quad + \int_{\mathbb{R}^k} \log \left(\exp \left\{ - \frac{u}{2\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^* - \mathbf{x}_t \mathbf{W}^{*T}) (\mathbf{y}_t - \boldsymbol{\mu}^* - \mathbf{x}_t \mathbf{W}^{*T})^T - \frac{u}{2} \mathbf{x}_t \mathbf{x}_t^T \right\} \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u \\
 &\quad - \frac{u}{2} \int_{\mathbb{R}^k} \left(\frac{1}{\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^* - \mathbf{x}_t \mathbf{W}^{*T}) (\mathbf{y}_t - \boldsymbol{\mu}^* - \mathbf{x}_t \mathbf{W}^{*T})^T + \mathbf{x}_t \mathbf{x}_t^T \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u + \log \pi_{U | \Psi^*}(u) \\
 &\quad - \frac{u}{2} \int_{\mathbb{R}^k} \left(\frac{1}{\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^*) (\mathbf{y}_t - \boldsymbol{\mu}^*)^T - \frac{2}{\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \mathbf{x}_t^T + \mathbf{x}_t \left(\frac{1}{\sigma^{*2}} \mathbf{W}^{*T} \mathbf{W}^* + \mathbb{I}_k \right) \mathbf{x}_t^T \right) \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u - \frac{u}{2\sigma^{*2}} \text{Tr} \left\{ (\mathbf{y}_t - \boldsymbol{\mu}^*)^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \right\} \\
 &\quad - \frac{u}{2} \left(\frac{1}{\sigma^{*2}} \int_{\mathbb{R}^k} \text{Tr} \left\{ \mathbf{x}_t^T \mathbf{x}_t (\mathbf{W}^{*T} \mathbf{W}^* + \sigma^{*2} \mathbb{I}_k) \right\} \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t - \frac{2}{\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \int_{\mathbb{R}^k} \mathbf{x}_t^T \varphi(\mathbf{x}_t; \tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}) d\mathbf{x}_t \right) \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u + \log \pi_{U | \Psi^*}(u) \\
 &\quad - \frac{u}{2} \left(\frac{1}{\sigma^{*2}} \text{Tr} \left\{ (\mathbf{y}_t - \boldsymbol{\mu}^*)^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \right\} - \frac{2}{\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \tilde{\boldsymbol{\mu}}_t^T + \frac{1}{\sigma^{*2}} \text{Tr} \left\{ (\tilde{\boldsymbol{\Sigma}} + \tilde{\boldsymbol{\mu}}_t^T \tilde{\boldsymbol{\mu}}_t) \mathbf{M}^* \right\} \right) \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u - \frac{u}{2\sigma^{*2}} \text{Tr} \left\{ (\mathbf{y}_t - \boldsymbol{\mu}^*)^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \right\} \\
 &\quad - \frac{u}{2} \left(\frac{1}{\sigma^{*2}} \text{Tr} \left\{ \frac{\sigma^2}{u} \mathbf{M}^{-1} \mathbf{M}^* + \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}) \mathbf{W} \mathbf{M}^{-1} \mathbf{M}^* \right\} - \frac{2}{\sigma^{*2}} (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{y}_t - \boldsymbol{\mu})^T \right) \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u - \frac{u}{2\sigma^{*2}} \text{Tr} \left\{ (\mathbf{y}_t - \boldsymbol{\mu}^*)^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \right\} \\
 &\quad - \frac{u}{2\sigma^{*2}} \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}) \mathbf{W} \mathbf{M}^{-1} \mathbf{M}^* - 2(\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} - \frac{\sigma^2}{2\sigma^{*2}} \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{M}^* \right\}
 \end{aligned}$$

where $\mathbf{M}^* = \mathbf{W}^{*T}\mathbf{W}^* + \sigma^{*2}\mathbb{I}_k$. \square

Lemma 3. Let the d dimensional observation vector \mathbf{Y}_t be modelled as in Equation (1) with the assumptions given in Equation (4) for scalar random variable Gamma $U \sim \Gamma(\frac{v}{2}, \frac{v}{2})$. We define function $H : \mathbb{R}^{N \times d} \times \mathbb{R}_+ \rightarrow \mathbb{R}$, $H(\mathbf{y}_{1:N}, u; \Psi) := \prod_{t=1}^N h(\mathbf{y}_t, u; \Psi)$ being a product of functions $h : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}$ defined in Lemma 1 over N realisation of vector \mathbf{Y}_t , $\mathbf{y}_{1:N}$, in time $t = 1, \dots, N$. Let u be a single realisation of U . It is true that

$$\pi_{U|\Psi}(u) \cdot H(\mathbf{y}_{1:N}, u; \Psi) = C_H(\mathbf{y}_{1:N}; \Psi) \cdot \tilde{\pi}_{U|\Psi}(u)$$

where $\tilde{\pi}_{U|\Psi}(u)$ is a density function of a Gamma random variable $U \sim \Gamma(\alpha, \beta(\mathbf{y}_{1:N}; \Psi))$ for $\alpha = \frac{v}{2} + \frac{dN}{2}$ and $\beta, C_H : \mathbb{R}^{N \times d} \rightarrow \mathbb{R}$ being defined as

$$\begin{aligned} \beta(\mathbf{y}_{1:N}; \Psi) &:= \frac{v}{2} + \frac{1}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \\ C_H(\mathbf{y}_{1:N}; \Psi) &:= (2\pi)^{-\frac{dN}{2}} (\sigma^2)^{-N\frac{d-k}{2}} \left| \mathbf{M} \right|^{-\frac{N}{2}} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}} \Gamma(\alpha)}{\Gamma\left(\frac{v}{2}\right)} \beta(\mathbf{y}_{1:N}; \Psi)^{-\alpha}. \end{aligned}$$

Proof. Let us recall the notation and the definition of the function h from Lemma 1, that is

$$h(\mathbf{y}_t, u; \Psi) = (2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d-k}{2}} u^{\frac{d}{2}} \left| \mathbf{M} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{u}{2\sigma^2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\}$$

The function $H : \mathbb{R}^{N \times d} \times \mathbb{R}_+ \rightarrow \mathbb{R}$, $H(\mathbf{y}_{1:N}, u; \Psi) := \prod_{t=1}^N h(\mathbf{y}_t, u; \Psi)$, is calculated as follows

$$\begin{aligned} H(\mathbf{y}_{1:N}, u; \Psi) &= \prod_{t=1}^N \left((2\pi)^{-\frac{d}{2}} (\sigma^2)^{-\frac{d-k}{2}} u^{\frac{d}{2}} \left| \mathbf{M} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{u}{2\sigma^2} (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\} \right) \\ &= (2\pi)^{-\frac{dN}{2}} (\sigma^2)^{-N\frac{d-k}{2}} u^{\frac{dN}{2}} \left| \mathbf{M} \right|^{-\frac{N}{2}} \exp \left\{ -\frac{u}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\} \end{aligned}$$

We show in the next part that given $U \sim \Gamma\left(\frac{v}{2}, \frac{v}{2}\right)$, the expression $\pi_{U|\Psi}(u) \cdot H(\mathbf{y}_{1:N}, u; \Psi)$ is proportional to a gamma distribution, that is

$$\begin{aligned} \pi_{U|\Psi}(u) \cdot H(\mathbf{y}_{1:N}, u; \Psi) &= (2\pi)^{-\frac{dN}{2}} (\sigma^2)^{-N\frac{d-k}{2}} \left| \mathbf{M} \right|^{-\frac{N}{2}} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \\ &\quad \times u^{\frac{v}{2}-1} e^{-\frac{v}{2}u} u^{\frac{dN}{2}} \exp \left\{ -\frac{u}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right\} \\ &\propto u^{\frac{v}{2} + \frac{dN}{2} - 1} \exp \left\{ -u \left(\frac{v}{2} + \frac{1}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T \right) \right\} \\ &\propto \frac{\beta(\mathbf{y}_{1:N}; \Psi)^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-u\beta(\mathbf{y}_{1:N}; \Psi)} =: \tilde{\pi}_{U|\Psi}(u) \end{aligned}$$

for $\alpha = \frac{v}{2} + \frac{dN}{2}$ and $\beta : \mathbb{R}^{N \times d} \rightarrow \mathbb{R}_+$ being a function defined as

$$\beta(\mathbf{y}_{1:N}; \Psi) = \frac{v}{2} + \frac{1}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t - \boldsymbol{\mu}) \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) (\mathbf{y}_t - \boldsymbol{\mu})^T.$$

Recall that the relation of proportionality means that we can find C_H which is not a function of u , such that

$$\pi_{U|\Psi}(u) \cdot H(\mathbf{y}_{1:N}, u; \Psi) = C_H \cdot \tilde{\pi}_{U|\Psi}(u).$$

In our case, C_H is a function of $\mathbf{y}_{1:N}$ as

$$C_H(\mathbf{y}_{1:N}; \Psi) = \frac{\pi_{U|\Psi}(u)H(\mathbf{y}_{1:N}, u; \Psi)}{\tilde{\pi}_{U|\Psi}(u)} = (2\pi)^{-\frac{dN}{2}}(\sigma^2)^{-N\frac{d-k}{2}}|\mathbf{M}|^{-\frac{N}{2}}\frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}\Gamma(\alpha)}{\Gamma\left(\frac{v}{2}\right)}\beta(\mathbf{y}_{1:N}; \Psi)^{-\alpha}$$

□

Lemma 4. Let a d dimensional observation vector \mathbf{Y}_t be modelled as in Equation (1) with assumptions given in Equation (4) for a scalar random variable $U \sim \Gamma\left(\frac{v}{2}, \frac{v}{2}\right)$. We define function $H : \mathbb{R}^{N \times d} \times \mathbb{R}_+ \rightarrow \mathbb{R}$, $H(\mathbf{y}_{1:N}, u; \Psi) := \prod_{s=1}^N h(\mathbf{y}_s, u; \Psi)$ being a product of functions $h : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}$ defined in Lemma 1 over N realisation of vector \mathbf{Y}_t , $\mathbf{y}_{1:N}$, in time $t = 1, \dots, N$. Let u be a single realisation of U . It is true that there exists a scalar $C_H(\Psi) \in \mathbb{R}$ equal to $C_H(\Psi) = (\sigma^2)^{\frac{kN}{2}}|\mathbf{M}|^{-\frac{N}{2}}\left|\left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\right)\right|^{-\frac{N}{2}}$ such that

$$H(\mathbf{y}_{1:N}, u; \Psi) = C_H(\Psi) \cdot \prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t|U, \Psi}(\mathbf{y}_t)$$

where $\tilde{\pi}_{\mathbf{Y}_t|U, \Psi}(\mathbf{y}_t) := \phi(\mathbf{y}_t, \boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}})$ for $\tilde{\boldsymbol{\Sigma}} = \frac{\sigma^2}{u}(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T)^{-1}$ and $\phi(\mathbf{y}, \boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}})$ being probability density function of d -dimensional Gaussian random vector with mean $\boldsymbol{\mu}$ and a covariance matrix $\tilde{\boldsymbol{\Sigma}}$.

Proof. Given the definition of function h from Lemma 1

$$h(\mathbf{y}_t, u; \Psi) = (2\pi)^{-\frac{d}{2}}(\sigma^2)^{-\frac{d-k}{2}}u^{\frac{d}{2}}|\mathbf{M}|^{-\frac{1}{2}}\exp\left\{-\frac{u}{2\sigma^2}(\mathbf{y}_t - \boldsymbol{\mu})\left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\right)(\mathbf{y}_t - \boldsymbol{\mu})^T\right\},$$

the function $H : \mathbb{R}^{N \times d} \times \mathbb{R}_+ \rightarrow \mathbb{R}$, which equals to $H(\mathbf{y}_{1:N}, u; \Psi) := \prod_{s=1}^N h(\mathbf{y}_s, u; \Psi)$, is calculated as follows

$$H(\mathbf{y}_{1:N}, u; \Psi) = \prod_{s=1}^N \left((2\pi)^{-\frac{d}{2}}(\sigma^2)^{-\frac{d-k}{2}}u^{\frac{d}{2}}|\mathbf{M}|^{-\frac{1}{2}}\exp\left\{-\frac{u}{2\sigma^2}(\mathbf{y}_t - \boldsymbol{\mu})\left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\right)(\mathbf{y}_t - \boldsymbol{\mu})^T\right\} \right)$$

Let us denote $\tilde{\boldsymbol{\Sigma}} = \frac{\sigma^2}{u}(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T)^{-1}$. Then

$$\begin{aligned} H(\mathbf{y}_{1:N}, u; \Psi) &= \prod_{s=1}^N \left((2\pi)^{-\frac{d}{2}}(\sigma^2)^{-\frac{d-k}{2}}u^{\frac{d}{2}}|\mathbf{M}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu})\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_t - \boldsymbol{\mu})^T\right\} \right) \\ &= \prod_{s=1}^N \left((2\pi)^{-\frac{d}{2}}(\sigma^2)^{-\frac{d-k}{2}}u^{\frac{d}{2}}|\mathbf{M}|^{-\frac{1}{2}}|\tilde{\boldsymbol{\Sigma}}|^{\frac{1}{2}}|\tilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu})\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_t - \boldsymbol{\mu})^T\right\} \right) \\ &= \prod_{s=1}^N \left((\sigma^2)^{-\frac{d-k}{2}}u^{\frac{d}{2}}|\mathbf{M}|^{-\frac{1}{2}}|\tilde{\boldsymbol{\Sigma}}|^{\frac{1}{2}} \underbrace{(2\pi)^{-\frac{d}{2}}|\tilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu})\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_t - \boldsymbol{\mu})^T\right\}}_{\phi(\mathbf{y}_t; \boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}})} \right) \\ &= \prod_{s=1}^N \left((\sigma^2)^{-\frac{d-k}{2}}u^{\frac{d}{2}}|\mathbf{M}|^{-\frac{1}{2}}\left|\frac{\sigma^2}{u}(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T)^{-1}\right|^{\frac{1}{2}}\phi(\mathbf{y}_t; \boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}}) \right) \\ &= \prod_{t=1}^N \left((\sigma^2)^{\frac{k}{2}}|\mathbf{M}|^{-\frac{1}{2}}\left|\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\right|^{-\frac{1}{2}}\phi(\mathbf{y}_t; \boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}}) \right) \end{aligned}$$

By denoting $C_H(\Psi) = (\sigma^2)^{\frac{kN}{2}} |\mathbf{M}|^{-\frac{N}{2}} \left| \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T \right) \right|^{-\frac{N}{2}}$ and $\tilde{\pi}_{\mathbf{Y}_t|U,\Psi}(\mathbf{y}_t) := \phi(\mathbf{y}_t; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ we have that the function H can be decomposed into a scalar and product of Gaussian, multivariate densities with the same mean and covariance

$$H(\mathbf{y}_{1:N}, u; \Psi) = C_H(\Psi) \cdot \prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t|U,\Psi}(\mathbf{y}_t)$$

□

Lemma 5. Let a d dimensional observation vector \mathbf{Y}_t has a probability density function $\tilde{\pi}_{\mathbf{Y}_t|U,\Psi}(\mathbf{y}_t)$ specified in Lemma 4 and let U be specified as in Section 3.2. Given the partition of the vector, $\mathbf{Y}_t = [\mathbf{Y}_t^o, \mathbf{Y}_t^m]$, the conditional distribution of $\mathbf{Y}_t|\mathbf{Y}_t^o, U, \Psi$ is the following

$$\mathbf{Y}_t|\mathbf{Y}_t^o, U, \Psi \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix}, \frac{\sigma^2}{U} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{mm}^{-1} - \mathbf{V}_{mo}^{-1} \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix} \right)$$

for $\mathbf{V} = \mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T$, where \mathbf{V}_{oo} and \mathbf{V}_{mm} or \mathbf{V}_{oo}^{-1} and \mathbf{V}_{mm}^{-1} are $d_o \times d_o$ and $d_m \times d_m$ square, symmetric submatrices of \mathbf{V} or \mathbf{V}^{-1} with rows (the left side lower index) and columns (the right side lower index) corresponding to observed entries and missing entries of the vector \mathbf{Y}_t , respectively. The matrices $\mathbf{V}_{om} = \mathbf{V}_{mo}^T$ or $\mathbf{V}_{om}^{-1} = \mathbf{V}_{mo}^{-1T}$ is a submatrix of the matrix \mathbf{V} or \mathbf{V}^{-1} , respectively, which rows correspond to observed entries and columns to missing entries of the vector \mathbf{Y}_t .

Then we have the following solution to the integration problem

$$\begin{aligned} \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) &= \int_{\mathbb{R}^{d_m}} w(\mathbf{y}_t, u; \Psi, \Psi^*) \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \\ &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u - \frac{\sigma^2}{2\sigma^{*2}} \text{Tr} \{ \mathbf{M}^{-1} \mathbf{M}^* \} \\ &\quad - \frac{u}{2\sigma^{*2}} \left(\text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - 2\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \boldsymbol{\mu}^{*T} + \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T} \right\} \right) \\ &\quad - 2 \text{Tr} \left\{ \left(\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] (\boldsymbol{\mu} + \boldsymbol{\mu}^*)^T + \boldsymbol{\mu}^* \boldsymbol{\mu}^T \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \\ &\quad + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - 2\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \boldsymbol{\mu}^T + \boldsymbol{\mu} \boldsymbol{\mu}^T \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{M}^* \right\} \end{aligned}$$

for the corresponding moments of the conditional distribution

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] &= \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix}, \\ \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t^T \mathbf{Y}_t] &= \frac{\sigma^2}{U} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{mm}^{-1} - \mathbf{V}_{mo}^{-1} \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix} + \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]. \end{aligned}$$

Proof. Let us denote $\tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) := \int_{\mathbb{R}^{d_m}} w(\mathbf{y}_t, u; \Psi, \Psi^*) \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m$. Using the definition of the function w from Lemma 2, we have that

$$\tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) = \int_{\mathbb{R}^{d_m}} w(\mathbf{y}_t, u; \Psi, \Psi^*) \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m$$

$$\begin{aligned}
 &= \int_{\mathbb{R}^{d_m}} \left(-\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u \right) \tilde{\pi}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \\
 &\quad - \frac{u}{2\sigma^{*2}} \int_{\mathbb{R}^{d_m}} \text{Tr} \left\{ (\mathbf{y}_t - \boldsymbol{\mu}^*)^T (\mathbf{y}_t - \boldsymbol{\mu}^*) - 2(\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}^*) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \tilde{\pi}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \\
 &\quad - \frac{u}{2\sigma^{*2}} \int_{\mathbb{R}^{d_m}} \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{y}_t - \boldsymbol{\mu})^T (\mathbf{y}_t - \boldsymbol{\mu}) \mathbf{W} \mathbf{M}^{-1} \mathbf{M}^* \right\} \tilde{\pi}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \\
 &\quad - \frac{\sigma^2}{2\sigma^{*2}} \int_{\mathbb{R}^{d_m}} \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{M}^* \right\} \tilde{\pi}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \\
 &= -\frac{d+k}{2} \log(2\pi) - \frac{d}{2} \log(\sigma^{*2}) + \frac{d+k}{2} \log u - \frac{\sigma^2}{2\sigma^{*2}} \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{M}^* \right\} \\
 &\quad - \frac{u}{2\sigma^{*2}} \left(\text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - 2\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \boldsymbol{\mu}^{*T} + \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T} \right\} \right. \\
 &\quad \left. - 2 \text{Tr} \left\{ \left(\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] (\boldsymbol{\mu} + \boldsymbol{\mu}^*)^T + \boldsymbol{\mu}^* \boldsymbol{\mu}^T \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \right. \\
 &\quad \left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - 2\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \boldsymbol{\mu}^T + \boldsymbol{\mu} \boldsymbol{\mu}^T \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{M}^* \right\} \right)
 \end{aligned}$$

Let us recall the fact $\tilde{\pi}_{\mathbf{Y}_t | U, \Psi}(\mathbf{y}_t)$ is a probability density function of multivariate Gaussian random variable, that is, $\mathbf{Y}_t = [\mathbf{Y}_t^o, \mathbf{Y}_t^m]$ conditioned on U is a multivariate Gaussian random vector with the mean $\boldsymbol{\mu}$ and the covariance matrix $\tilde{\boldsymbol{\Sigma}}$.

We denote $\mathbf{V} := \mathbb{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^T$ such that $\tilde{\boldsymbol{\Sigma}} = \frac{\sigma^2}{u} \mathbf{V}^{-1}$. The distribution of a subvector of Gaussian random variable conditioned of the other subvector of the same random variable is widely known, we may refer to [Gupta and Nagar \(1999\)](#) for further details. Hence, the distribution of $\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi$ is multivariate Gaussian, that is

$$\mathbf{Y}_t | \mathbf{Y}_t^o, U, \Psi \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix}, \frac{\sigma^2}{u} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{mm}^{-1} - \mathbf{V}_{mo}^{-1} \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix} \right)$$

where \mathbf{V}_{oo} and \mathbf{V}_{mm} or \mathbf{V}_{oo}^{-1} and \mathbf{V}_{mm}^{-1} are $d_o \times d_o$ and $d_m \times d_m$ square, symmetric submatrices of \mathbf{V} or \mathbf{V}^{-1} with rows (the left side lower index) and columns (the right side lower index) corresponding to observed entries and missing entries of the vector \mathbf{Y}_t , respectively. The matrices $\mathbf{V}_{om} = \mathbf{V}_{mo}^T$ or $\mathbf{V}_{om}^{-1} = \mathbf{V}_{mo}^{-1T}$ is a submatrix of the matrix \mathbf{V} or \mathbf{V}^{-1} , respectively, which rows correspond to observed entries and columns to missing entries of the vector \mathbf{Y}_t . Let us denote

$$\mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{mm}^{-1} - \mathbf{V}_{mo}^{-1} \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix}.$$

The required moments can be easily derived given the conditional distribution of $\mathbf{Y}_t | \mathbf{Y}_t^o, U, \Psi$, that is

$$\begin{aligned}
 \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] &= \begin{bmatrix} \mathbf{Y}_t^o \\ \boldsymbol{\mu}_m + (\mathbf{Y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo}^{-1} \mathbf{V}_{om} \end{bmatrix}, \\
 \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t^T \mathbf{Y}_t] &= \frac{\sigma^2}{u} \mathbf{Q} + \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t].
 \end{aligned}$$

□

Lemma 6. Let a d dimensional observation vector \mathbf{Y}_t has a probability density function $\tilde{\pi}_{\mathbf{Y}_t|U,\Psi}(\mathbf{y}_t)$ specified in Lemma 4 and let U be specified as in Section 3.2 . Given the partition of the vector , $\mathbf{Y}_t = [\mathbf{Y}_t^o, \mathbf{Y}_t^m]$, the marginal distribution of $\mathbf{Y}_t^o|U, \Psi$ is the following

$$\mathbf{Y}_t^o|U, \Psi \sim \mathcal{N}\left(\boldsymbol{\mu}_o, \frac{\sigma^2}{U} \mathbf{V}_{oo}^{-1}\right)$$

where $\boldsymbol{\mu}_o$ is a subvector of $\boldsymbol{\mu}$ which elements correspond to the vector with the observed entries \mathbf{Y}_o and \mathbf{V}_{oo}^{-1} is a square, symmetric submatrix of $\mathbf{V}^{-1} = \left(\mathbb{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\right)^{-1}$ which correspond to the vector with the observed entries by rows and columns.

Let $\mathbf{y}_{1:N}^o$ be N realisations of vector \mathbf{Y}_t^o . We can specify the function $C_\beta : \mathbb{R}^{d_o \times N} \rightarrow \mathbb{R}$

$$C_\beta(\mathbf{y}_{1:N}^o, \Psi) = \frac{(2\pi)^{-\frac{d_o N}{2}} (\sigma^2)^{-\frac{d_o N}{2}} |\mathbf{V}_{oo}|^{\frac{N}{2}} \left(\frac{v}{2}\right)^{\frac{v}{2}} \Gamma(\alpha)}{\Gamma\left(\frac{v}{2}\right)} \beta(\mathbf{y}_{1:N}^o; \Psi)^{-\alpha},$$

which is constant with respect to U , such that

$$\pi_{U|\Psi}(u) \cdot \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o)\right) = C_\beta(\mathbf{y}_{1:N}^o, \Psi) \cdot \tilde{\pi}_{U|\Psi}(u)$$

where $\tilde{\pi}_{U|\Psi}(u)$ is a density of univariate Gamma random variable $U \sim \Gamma(\alpha, \beta(\mathbf{y}_{1:N}^o; \Psi))$ for the scalar $\alpha = \frac{v+d_o N}{2}$ and a function $\beta : \mathbb{R}^{d_o \times N} \rightarrow \mathbb{R}$

$$\beta(\mathbf{y}_{1:N}^o; \Psi) = \frac{v}{2} + \frac{1}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} (\mathbf{y}_t^o - \boldsymbol{\mu}_o)^T.$$

Proof. We show that $\pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o)\right) \propto \tilde{\pi}_{U|\Psi}(u)$, where $\tilde{\pi}_{U|\Psi}(u)$ is a density of univariate Gamma random variable. The density $\tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o)$ is a probability function of \mathbf{Y}_t^o which is multivariate Gaussian with the mean $\boldsymbol{\mu}_o$ and the covariance matrix $\boldsymbol{\Sigma}_{oo} = \frac{\sigma^2}{u} \mathbf{V}_{oo}^{-1}$. The reader may refer to Gupta and Nagar (1999) to find the derivation of a marginal distribution of a subvector of multivariate Gaussian random variable. Using this fact, we have

$$\begin{aligned} \pi_{U|\Psi}(u) \cdot \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o)\right) &= \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} u^{\frac{v}{2}-1} e^{-\frac{v}{2}u} \\ &\times \prod_{t=1}^N (2\pi)^{-\frac{d}{2}} \left|\frac{\sigma^2}{u} \mathbf{V}_{oo}^{-1}\right|^{-\frac{1}{2}} \exp\left\{-\frac{u}{2\sigma^2} (\mathbf{y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} (\mathbf{y}_t^o - \boldsymbol{\mu}_o)^T\right\} \\ &= \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} u^{\frac{v}{2}-1} e^{-\frac{v}{2}u} (2\pi)^{-\frac{d_o N}{2}} (\sigma^2)^{-\frac{d_o N}{2}} u^{\frac{d_o N}{2}} |\mathbf{V}_{oo}|^{\frac{N}{2}} \exp\left\{-\frac{u}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} (\mathbf{y}_t^o - \boldsymbol{\mu}_o)^T\right\} \\ &= (2\pi)^{-\frac{d_o N}{2}} (\sigma^2)^{-\frac{d_o N}{2}} |\mathbf{V}_{oo}|^{\frac{N}{2}} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} u^{\frac{v+d_o N}{2}-1} \exp\left\{-u\left(\frac{v}{2} + \frac{1}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} (\mathbf{y}_t^o - \boldsymbol{\mu}_o)^T\right)\right\} \end{aligned}$$

Let us denote the scalar $\alpha = \frac{v+d_o N}{2}$ and define a function $\beta : \mathbb{R}^{d_o \times N} \rightarrow \mathbb{R}$ such that

$$\beta(\mathbf{y}_{1:N}^o; \Psi) = \frac{v}{2} + \frac{1}{2\sigma^2} \sum_{t=1}^N (\mathbf{y}_t^o - \boldsymbol{\mu}_o) \mathbf{V}_{oo} (\mathbf{y}_t^o - \boldsymbol{\mu}_o)^T.$$

We obtain the following proportionality

$$\pi_{U|\Psi}(u) \cdot \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \propto u^{\alpha-1} \exp \left\{ -u\beta(\mathbf{y}_{1:N}^o; \Psi) \right\} \propto \tilde{\pi}_{U|\Psi}(u),$$

where $\tilde{\pi}_{U|\Psi}(u)$ is a probability density function of the a univariate random variable $U \sim \Gamma(\alpha, \beta(\mathbf{y}_{1:N}^o; \Psi))$. We can specify the function $C_\beta : \mathbb{R}^{d_o \times N} \rightarrow \mathbb{R}$, which is a constant with respect to U , such that

$$\pi_{U|\Psi}(u) \cdot \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) = C_\beta(\mathbf{y}_{1:N}^o, \Psi) \cdot \tilde{\pi}_{U|\Psi}(u)$$

After some derivation, we obtain that

$$\begin{aligned} C_\beta(\mathbf{y}_{1:N}^o, \Psi) &= \frac{\pi_{U|\Psi}(u) \cdot \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right)}{\tilde{\pi}_{U|\Psi}(u)} = \frac{(2\pi)^{-\frac{d_o N}{2}} (\sigma^2)^{-\frac{d_o N}{2}} \left| \mathbf{V}_{oo} \right|^{\frac{N}{2}} \left(\frac{v}{2} \right)^{\frac{v}{2}}}{\frac{\beta(\mathbf{y}_{1:N}^o; \Psi)^\alpha}{\Gamma(\alpha)}} \\ &= \frac{(2\pi)^{-\frac{d_o N}{2}} (\sigma^2)^{-\frac{d_o N}{2}} \left| \mathbf{V}_{oo} \right|^{\frac{N}{2}} \left(\frac{v}{2} \right)^{\frac{v}{2}} \Gamma(\alpha)}{\Gamma\left(\frac{v}{2}\right)} \beta(\mathbf{y}_{1:N}^o; \Psi)^{-\alpha} \end{aligned}$$

□

D.3.2. The Proof of Theorem 3

Proof. We follow the similar steps as in the proof of Theorem 1 but utilising the loglikelihood from Equation (5) and corresponding assumptions. Hence, the E-step of EM algorithm for Identical and Conditionally Independent t-Student with missing values is calculated as

$$\begin{aligned} \tilde{Q}(\Psi, \Psi^*) &= \mathbb{E}_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N}, U | \mathbf{Y}_{1:N}^o, \Psi} \left[\log \pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N}, U | \Psi^*}(\mathbf{Y}_{1:N}, \mathbf{X}_{1:N}, U) \right] \\ &= \int_{\mathbb{R}_+} \int_{\mathbb{R}^{N \times d_m}} \int_{\mathbb{R}^{N \times k}} \left\{ \log \left(\pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N}, U | \Psi^*}(\mathbf{y}_{1:N}, \mathbf{x}_{1:N}, u) \right) \right. \\ &\quad \times \left. \frac{\pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N}, U | \Psi}(\mathbf{y}_{1:N}, \mathbf{x}_{1:N}, u)}{\pi_{\mathbf{Y}_{1:N}^o | \Psi}(\mathbf{y}_{1:N}^o)} \right\} d\mathbf{x}_{1:N} d\mathbf{y}_{1:N}^m du \\ &= \frac{1}{\pi_{\mathbf{Y}_{1:N}^o | \Psi}(\mathbf{y}_{1:N}^o)} \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \int_{\mathbb{R}^{N \times d_m}} \int_{\mathbb{R}^{N \times k}} \left\{ \log \left(\pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N}, U | \Psi^*}(\mathbf{y}_{1:N}, \mathbf{x}_{1:N}, u) \right) \right. \\ &\quad \times \left. \pi_{\mathbf{Y}_{1:N} | \mathbf{X}_{1:N}, U, \Psi}(\mathbf{y}_{1:N}) \pi_{\mathbf{X}_{1:N} | U, \Psi}(\mathbf{x}_{1:N}) \right\} d\mathbf{x}_{1:N} d\mathbf{y}_{1:N}^m du \xrightarrow{\text{applying Lemma 1 and 2}} \\ &= \frac{1}{\pi_{\mathbf{Y}_{1:N}^o | \Psi}(\mathbf{y}_{1:N}^o)} \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \int_{\mathbb{R}^{N \times d_m}} \prod_{s=1}^N h(\mathbf{y}_s, u; \Psi) \left(\log \pi_{U|\Psi^*}(u) + \sum_{t=1}^N w(\mathbf{y}_t, u; \Psi, \Psi^*) \right) d\mathbf{y}_{1:N}^m du \\ &= \frac{1}{\pi_{\mathbf{Y}_{1:N}^o | \Psi}(\mathbf{y}_{1:N}^o)} \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left\{ \int_{\mathbb{R}^{N \times d_m}} C_H(\Psi) \cdot \left(\prod_{s=1}^N \tilde{\pi}_{\mathbf{Y}_s | U, \Psi}(\mathbf{y}_s) \right) \log \pi_{U|\Psi^*}(u) \mathbf{y}_{1:N}^m \right. \\ &\quad \left. + \int_{\mathbb{R}^{N \times d_m}} C_H(\Psi) \cdot \left(\prod_{s=1}^N \tilde{\pi}_{\mathbf{Y}_s | U, \Psi}(\mathbf{y}_s) \right) \left(\sum_{t=1}^N w(\mathbf{y}_t, u; \Psi, \Psi^*) \right) d\mathbf{y}_{1:N}^m \right\} du \end{aligned}$$

where $C_H(\Psi)$ and $\tilde{\pi}_{\mathbf{Y}_t|U,\Psi}(\mathbf{y}_t)$ are specified in Lemma 4 such that

$$\prod_{s=1}^N h(\mathbf{y}_s, u; \Psi) = C_H(\Psi) \prod_{s=1}^N \tilde{\pi}_{\mathbf{Y}_s|U,\Psi}(\mathbf{y}_s)$$

Given the partition of the observation vector \mathbf{Y}_t into the vector of observed and missing entries in time t , \mathbf{Y}_t^o and \mathbf{Y}_t^m , respectively, that is $\mathbf{Y}_t = [\mathbf{Y}_t^o, \mathbf{Y}_t^m]$ and applying Chain rule of probability, we obtain the following equality

$$\tilde{\pi}_{\mathbf{Y}_t|U,\Psi}(\mathbf{y}_t) = \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o,U,\Psi}(\mathbf{y}_t^m) \cdot \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o)$$

we have that

$$\begin{aligned} \tilde{Q}(\Psi, \Psi^*) &= \frac{C_H(\Psi)}{\pi_{\mathbf{Y}_{1:N}^o|\Psi}(\mathbf{y}_{1:N}^o)} \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \int_{\mathbb{R}^{N \times d_m}} \left\{ \left(\log \pi_{U|\Psi^*}(u) + \sum_{t=1}^N w(\mathbf{y}_t, u; \Psi, \Psi^*) \right) \right. \\ &\quad \left. \times \prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o,U,\Psi}(\mathbf{y}_t^m) \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right\} d\mathbf{y}_{1:N}^m du \\ &= \frac{C_H(\Psi)}{\pi_{\mathbf{Y}_{1:N}^o|\Psi}(\mathbf{y}_{1:N}^o)} \left\{ \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \log \pi_{U|\Psi^*}(u) \underbrace{\int_{\mathbb{R}^{N \times d_m}} \prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o,U,\Psi}(\mathbf{y}_t^m) d\mathbf{y}_{1:N}^m}_{=1} du \right. \\ &\quad \left. + \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \int_{\mathbb{R}^{N \times d_m}} \left(\sum_{t=1}^N w(\mathbf{y}_t, u; \Psi, \Psi^*) \right) \prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o,U,\Psi}(\mathbf{y}_t^m) d\mathbf{y}_{1:N}^m du \right\} \\ &= \frac{C_H(\Psi)}{\pi_{\mathbf{Y}_{1:N}^o|\Psi}(\mathbf{y}_{1:N}^o)} \left\{ \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \log \pi_{U|\Psi^*}(u) du \right. \\ &\quad \left. + \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \sum_{t=1}^N \underbrace{\left\{ \int_{\mathbb{R}^{d_m}} w(\mathbf{y}_t, u; \Psi, \Psi^*) \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o,U,\Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \right\}}_{\text{the } t \text{ element}} du \right. \\ &\quad \left. \times \prod_{s=1, s \neq t}^N \underbrace{\left\{ \int_{\mathbb{R}^{d_m}} \tilde{\pi}_{\mathbf{Y}_s^m|\mathbf{Y}_s^o,U,\Psi}(\mathbf{y}_s^m) d\mathbf{y}_s^m \right\}}_{=1 \text{ since a density}} du \right\} \\ &= \frac{C_H(\Psi)}{\pi_{\mathbf{Y}_{1:N}^o|\Psi}(\mathbf{y}_{1:N}^o)} \left\{ \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \log \pi_{U|\Psi^*}(u) du \right. \\ &\quad \left. + \int_{\mathbb{R}_+} \pi_{U|\Psi}(u) \left(\prod_{t=1}^N \tilde{\pi}_{\mathbf{Y}_t^o|U,\Psi}(\mathbf{y}_t^o) \right) \sum_{t=1}^N \left\{ \int_{\mathbb{R}^{d_m}} w(\mathbf{y}_t, u; \Psi, \Psi^*) \tilde{\pi}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o,U,\Psi}(\mathbf{y}_t^m) d\mathbf{y}_t^m \right\} du \right\} \end{aligned}$$

Applying Lemma 5 and 6, we obtain the following simplified form of the function \tilde{Q}

$$\tilde{Q}(\Psi, \Psi^*) = \frac{C_H(\Psi) C_\beta(\mathbf{y}_{1:N}^o, \Psi)}{\pi_{\mathbf{Y}_{1:N}^o|\Psi}(\mathbf{y}_{1:N}^o)} \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \left(\log \pi_{U|\Psi^*}(u) + \sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o; \Psi, \Psi^*) \right) du.$$

where $\tilde{w}(\mathbf{y}_t^o; \Psi, \Psi^*)$ is derived in Lemma 5 and the density function $\tilde{\pi}_{U|\Psi}(u)$ is given in Lemma 6. \square

D.3.3. The Proof of Theorem 4

Proof. The maximizers of \tilde{Q} with respect to the vector of static parameters $\Psi^* = [\mathbf{W}^*, \mu^*, \sigma^{*2}]$ be calculates as

$$\begin{aligned} \nabla_{\Psi^*} \tilde{Q}(\Psi, \Psi^*) &= \frac{\partial}{\partial \Psi^*} \tilde{C}(\mathbf{y}_{1:N}^o, \Psi) \left\{ \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \log \pi_{U|\Psi^*}(u) du + \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) du \right\} \\ &= \tilde{C}(\mathbf{y}_{1:N}^o, \Psi) \left\{ \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \underbrace{\frac{\partial}{\partial \Psi^*} \log \pi_{U|\Psi^*}(u)}_{= 0 \text{ as includes only } v^*} du + \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \frac{\partial}{\partial \Psi^*} \sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) du \right\} \\ &= \tilde{C}(\mathbf{y}_{1:N}^o, \Psi) \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \sum_{t=1}^N \frac{\partial}{\partial \Psi^*} \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) du \end{aligned}$$

for $\tilde{C}(\mathbf{y}_{1:N}^o, \Psi) = \frac{C_H(\Psi) C_\beta(\mathbf{y}_{1:N}^o, \Psi)}{\pi_{\mathbf{y}_{1:N}^o|\Psi}(\mathbf{y}_{1:N}^o)}$. We applied Leibniz integral rule to swap the order of differentiating and integrating as the function $f(\Psi^*, u) = \tilde{\pi}_{U|\Psi}(u) \left(\log \pi_{U|\Psi^*}(u) + \sum_{t=1}^N w(\mathbf{y}_t, u; \Psi, \Psi^*) \right)$ is a linear function of u . It is integrable with respect to u and differentiable at Ψ^* and its gradient $\nabla_{\Psi^*} f$ is bounded by a function of u . Next we calculate the gradient of $\sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*)$ with respect to $[\mathbf{W}^*, \mu^*, \sigma^{*2}]$, that is

$$\begin{aligned} \frac{\partial}{\partial \mu^*} \sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) &= \\ &= -\frac{u}{2\sigma^{*2}} \sum_{t=1}^N -2\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] + 2\mu^* + 2\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} - 2\mu \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\ &= \frac{u}{\sigma^{*2}} \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] \left(\mathbf{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) - \mu^* + \mu \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\ &= u \frac{N}{\sigma^{*2}} \underbrace{\left(\frac{1}{N} \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] \right)}_{\bar{\mu}} \left(\mathbf{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) - \mu^* + \mu \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\ &= u \frac{N}{\sigma^{*2}} \left(\bar{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \left(\mathbf{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) - \mu^* + \mu \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \mathbf{W}^*} \sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) &= \\ &= \sum_{t=1}^N \left\{ \frac{u}{\sigma^{*2}} \left(\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t \mathbf{Y}_t^T] - \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] (\mu + \mu^*)^T + \mu^*{}^T \mu \right) \mathbf{W} \mathbf{M}^{-1} \right. \\ &\quad - \frac{u}{\sigma^{*2}} \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \left(\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t \mathbf{Y}_t^T] - 2\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] \mu^T + \mu \mu^T \right) \mathbf{W} \mathbf{M}^{-1} \\ &\quad \left. - \frac{\sigma^2}{\sigma^{*2}} \mathbf{W}^* \mathbf{M}^{-1} \right\} \\ &= \frac{N}{\sigma^{*2}} \left(\frac{u}{N} \sum_{t=1}^N \left\{ \frac{\sigma^2}{u} \mathbf{Q} + \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t]{}^T \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] - \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] (\mu + \mu^*)^T + \mu^*{}^T \mu \right\} \mathbf{W} \mathbf{M}^{-1} \right. \\ &\quad \left. - \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \frac{u}{N} \sum_{t=1}^N \left\{ \frac{\sigma^2}{u} \mathbf{Q} + \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t]{}^T \mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] - 2\mathbb{E}_{\mathbf{Y}_t^m|\mathbf{Y}_t^o, U, \Psi}[\mathbf{Y}_t] \mu^T + \mu \mu^T \right\} \mathbf{W} \mathbf{M}^{-1} \right) \end{aligned}$$

$$\begin{aligned}
 & -\sigma^2 \mathbf{W}^* \mathbf{M}^{-1} \Big) \\
 = & \frac{N}{\sigma^{*2}} \left(\underbrace{\sigma^2 \frac{1}{N} \sum_{t=1}^N \mathbf{Q} \mathbf{M}^{-1} \mathbf{W}^T}_{\bar{\mathbf{Q}}} - \sigma^2 \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \frac{1}{N} \sum_{t=1}^N \mathbf{Q} \mathbf{W} \mathbf{M}^{-1} - \sigma^2 \mathbf{W}^* \mathbf{M}^{-1} \right. \\
 & + u \frac{1}{N} \sum_{t=1}^N \left\{ \underbrace{\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] - \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T (\boldsymbol{\mu} + \boldsymbol{\mu}^*) + \boldsymbol{\mu}^{*T} \boldsymbol{\mu}}_{\bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)}} \right\} \mathbf{W} \mathbf{M}^{-1} \\
 & \left. - u \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \frac{1}{N} \sum_{t=1}^N \left\{ \underbrace{\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] - 2 \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T \boldsymbol{\mu} + \boldsymbol{\mu}^T \boldsymbol{\mu}}_{\bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)}} \right\} \mathbf{W} \mathbf{M}^{-1} \right) \\
 = & \frac{N}{\sigma^{*2}} \left((\sigma^2 \bar{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} - \mathbf{W}^* \left(\mathbf{M}^{-1} \mathbf{W}^T (\sigma^2 \bar{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right) \right) \\
 \\
 \frac{\partial}{\partial \sigma^{*2}} \sum_{t=1}^N \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) = & -\frac{dN}{2\sigma^{*2}} + \frac{1}{2(\sigma^{*2})^2} \sum_{t=1}^N \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right. \\
 & + u \left(\text{Tr} \left\{ \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - 2 \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \boldsymbol{\mu}^{*T} + \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T} \right\} \right. \\
 & \left. - 2 \text{Tr} \left\{ \left(\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] (\boldsymbol{\mu} + \boldsymbol{\mu}^*)^T + \boldsymbol{\mu}^{*T} \boldsymbol{\mu} \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \right. \\
 & \left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t \mathbf{Y}_t^T] - 2 \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \boldsymbol{\mu}^T + \boldsymbol{\mu} \boldsymbol{\mu}^T \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right) \Big) \\
 = & -\frac{dN}{2\sigma^{*2}} + \frac{N}{2(\sigma^{*2})^2} \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} + \text{Tr} \left\{ \sigma^2 \bar{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right\} \right. \\
 & \left. - 2 \text{Tr} \left\{ \left(\sigma^2 \bar{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \right. \\
 & \left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right)
 \end{aligned}$$

for the following notation

$$\begin{aligned}
 \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) &= \frac{1}{N} \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t], \\
 \bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) &= \frac{1}{N} \sum_{t=1}^N \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t]^T \mathbb{E}_{\mathbf{Y}_t^m | \mathbf{Y}_t^o, U, \Psi} [\mathbf{Y}_t] \\
 \bar{\mathbf{Q}} &= \frac{1}{N} \sum_{t=1}^N \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{mm}^{-1} - \mathbf{V}_{mo}^{-1} \mathbf{V}_{oo} \mathbf{V}_{om}^{-1} \end{bmatrix}, \\
 \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) &= \bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) (\boldsymbol{\mu} + \boldsymbol{\mu}^*)^T + \boldsymbol{\mu}^{*T} \boldsymbol{\mu}, \\
 \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) &= \bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \boldsymbol{\mu}^T + \boldsymbol{\mu}^T \boldsymbol{\mu}, \\
 \bar{\mathbf{C}}_{\boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) &= \bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \boldsymbol{\mu}^{*T} + \boldsymbol{\mu}^{*T} \boldsymbol{\mu}^*.
 \end{aligned}$$

Hence, the maximizers of \tilde{Q} with respect to the vector of static parameters $\Psi^* = [\mathbf{W}^*, \boldsymbol{\mu}^*, \sigma^{*2}]$ are calculated by solving the set of equation $\nabla_{\Psi^*} \tilde{Q} = \mathbf{0}$ what results in the following

$$\begin{aligned} \frac{\tilde{Q}(\Psi, \Psi^*)}{\partial \boldsymbol{\mu}^*} &= \tilde{C}(\mathbf{y}_{1:N}^o, \Psi) \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \sum_{t=1}^N \frac{\partial}{\partial \boldsymbol{\mu}^*} \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) du = \mathbf{0} \\ \iff \frac{N}{\sigma^{*2}} \int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) &\left(\bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) (\mathbf{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T}) - \boldsymbol{\mu}^* + \boldsymbol{\mu}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T} \right) du = \mathbf{0} \\ \iff \left(\bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) (\mathbf{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T}) - \boldsymbol{\mu}^* + \boldsymbol{\mu}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T} \right) &\int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du = \mathbf{0} \\ \iff \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) (\mathbf{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T}) - \boldsymbol{\mu}^* + \boldsymbol{\mu}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T} &= \mathbf{0} \\ \iff \boldsymbol{\mu}^* = \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) (\mathbf{I}_d - \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T}) + \boldsymbol{\mu}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{*T} \end{aligned}$$

$$\begin{aligned} \frac{\tilde{Q}_{ts,iid}(\Psi, \Psi^*)}{\partial \mathbf{W}^*} &= \tilde{C}(\mathbf{y}_{1:N}^o, \Psi) \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \sum_{t=1}^N \frac{\partial}{\partial \mathbf{W}^*} \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) du = \mathbf{0} \\ \iff \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) &\left(\sigma^2 \tilde{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}\mathbf{M}^{-1} du \\ &- \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \mathbf{W}^* \left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \tilde{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}\mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right) du = \mathbf{0} \\ \iff \left(\sigma^2 \tilde{\mathbf{Q}} + \underbrace{\int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du}_{A(\mathbf{y}_{1:N}^o, \Psi, \Psi^*)} \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) &\mathbf{W}\mathbf{M}^{-1} \\ &- \mathbf{W}^* \left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \tilde{\mathbf{Q}} + \underbrace{\int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du}_{A(\mathbf{y}_{1:N}^o, \Psi, \Psi^*)} \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}\mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right) = \mathbf{0} \\ \iff \left(\sigma^2 \tilde{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o, \Psi, \Psi^*) \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) &\mathbf{W}\mathbf{M}^{-1} \\ &- \mathbf{W}^* \left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \tilde{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o, \Psi, \Psi^*) \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}\mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right) = \mathbf{0} \\ \iff \mathbf{W}^* = \left(\sigma^2 \tilde{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o, \Psi, \Psi^*) \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) &\mathbf{W}\mathbf{M}^{-1} \\ \times \left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \tilde{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o, \Psi, \Psi^*) \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}\mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right)^{-1} \end{aligned}$$

$$\begin{aligned} \frac{\tilde{Q}(\Psi, \Psi^*)}{\partial \sigma^{*2}} &= \tilde{C}(\mathbf{y}_{1:N}^o, \Psi) \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) \sum_{t=1}^N \frac{\partial}{\partial \sigma^{*2}} \tilde{w}(\mathbf{y}_t^o, u; \Psi, \Psi^*) du = \mathbf{0} \\ \iff \int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) &\left\{ -\frac{dN}{2\sigma^{*2}} + \frac{N}{2(\sigma^{*2})^2} \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} + \text{Tr} \left\{ \sigma^2 \tilde{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right\} \right) \right. \\ &- 2 \text{Tr} \left\{ \left(\sigma^2 \tilde{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \\ &\left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \tilde{\mathbf{Q}} + u \bar{\mathbf{C}}_{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}\mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right\} du = \mathbf{0} \end{aligned}$$

$$\begin{aligned}
 &\Leftrightarrow -\frac{d}{\sigma^{*2}} + \frac{1}{(\sigma^{*2})^2} \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right. \\
 &\quad \left. + \text{Tr} \left\{ \sigma^2 \bar{\mathbf{Q}} + \underbrace{\int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du}_{A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right\} \right. \\
 &\quad \left. - 2 \text{Tr} \left\{ \left(\sigma^2 \bar{\mathbf{Q}} + \underbrace{\int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du}_{A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \right. \\
 &\quad \left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + \underbrace{\int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du}_{A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right) = 0 \\
 &\Leftrightarrow \sigma^{*2} = \frac{1}{d} \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} + \text{Tr} \left\{ \sigma^2 \bar{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \bar{\mathbf{C}}_{\mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right\} \right. \\
 &\quad \left. - 2 \text{Tr} \left\{ \left(\sigma^2 \bar{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \bar{\mathbf{C}}_{\mu, \mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \right. \\
 &\quad \left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \bar{\mathbf{C}}_{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right)
 \end{aligned}$$

Given Lemma 6, $\tilde{\pi}_{U|\Psi}(u)$ is the density function of univariate Gamma random variable, $U \sim \Gamma(\alpha, \beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*))$ and therefore we used the fact $\int_{\mathbb{R}_+} \tilde{\pi}_{U|\Psi}(u) du = 1$. Its expectation can be easily derived, that is

$$A(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) = \int_{\mathbb{R}_+} u \tilde{\pi}_{U|\Psi}(u) du = \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)}$$

Therefore, the solution to the set of equations given by $\nabla_{\Psi^*} \bar{\mathbf{Q}} = \mathbf{0}$ are the following

$$\begin{aligned}
 \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \left(\mathbf{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) + \boldsymbol{\mu} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\
 \mathbf{W}^* &= \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \\
 &\quad \times \left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right)^{-1} \\
 \sigma^{*2} &= \frac{1}{d} \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} + \text{Tr} \left\{ \sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu^*}^m(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right\} \right. \\
 &\quad \left. - 2 \text{Tr} \left\{ \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \right. \\
 &\quad \left. + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu}^m(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \right)
 \end{aligned}$$

The reader may notice that the vector $\boldsymbol{\mu}^*$ is a linear function \mathbf{W}^* and vice versa. Hence, we can solve the following linear system of equations

$$\begin{aligned}
 \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \left(\mathbf{I}_d - \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \right) + \boldsymbol{\mu} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \\
 \mathbf{W}^* &= \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1}
 \end{aligned}$$

$$\times \underbrace{\left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right)^{-1}}_{I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1}}$$

by substituting the values of \mathbf{W}^* to the top equation by its values in the bottom equation, that is

$$\begin{aligned} \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) + (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} \\ &\times \left(\left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \right)^T \\ \iff \\ \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) + \sigma^2 (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{Q}} \\ &+ \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{M}^{-1} \mathbf{W}^T I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{W} \mathbf{M}^{-1} \bar{\mathbf{C}}_{\mu, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \end{aligned}$$

Next we substitute $\bar{\mathbf{C}}_{\mu, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) = \bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - (\boldsymbol{\mu} + \boldsymbol{\mu}^*)^T \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) + \boldsymbol{\mu}^{*T} \boldsymbol{\mu}$ which contains the vector $\boldsymbol{\mu}^*$. We obtain that

$$\begin{aligned} \boldsymbol{\mu}^* &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) + \sigma^2 (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{Q}} \\ &+ \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \\ &+ \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\boldsymbol{\mu}^{*T} \boldsymbol{\mu} - \boldsymbol{\mu}^{*T} \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \\ \iff \\ \boldsymbol{\mu}^* &- \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right)^T \boldsymbol{\mu}^* \\ &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) + \sigma^2 (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{Q}} \\ &+ \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \\ \iff \\ &\left(1 - \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right)^T \right) \boldsymbol{\mu}^* \\ &= \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) + \sigma^2 (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \bar{\mathbf{Q}} \\ &+ \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)) \mathbf{W} \mathbf{M}^{-1} I(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)^{-1} \mathbf{M}^{-1} \mathbf{W}^T \left(\bar{\mathbf{S}}(\mathbf{y}_{1:N}^o; \Psi) - \boldsymbol{\mu}^T \bar{\boldsymbol{\mu}}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \end{aligned}$$

for

$$\begin{aligned} \mathbf{W}^* &= \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \boldsymbol{\mu}^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \\ &\times \left(\mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} + \sigma^2 \mathbf{M}^{-1} \right)^{-1} \end{aligned}$$

$$\begin{aligned} \sigma^{*2} = & \frac{1}{d} \left(\sigma^2 \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} + \text{Tr} \left\{ \sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right\} \right) \\ & - 2 \text{Tr} \left\{ \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu, \mu^*}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W}^* \mathbf{M}^{-1} \mathbf{W}^T \right\} \\ & + \text{Tr} \left\{ \mathbf{M}^{-1} \mathbf{W}^T \left(\sigma^2 \bar{\mathbf{Q}} + \frac{\alpha}{\beta(\mathbf{y}_{1:N}^o; \Psi, \Psi^*)} \bar{\mathbf{C}}_{\mu}(\mathbf{y}_{1:N}^o; \Psi, \Psi^*) \right) \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^{*T} \mathbf{W}^* \right\} \end{aligned}$$

□

Appendix E. Synthetic Data Case Studies

In the next section we demonstrate the performance of the feature extraction frameworks introduced in Section 4, the two t-Student PPCA frameworks, as well as the standard Gaussian Probabilistic Principal Component Analysis methodology introduced by [Tipping and Bishop \(1999\)](#) and its robust alternative which adapts the robust estimation of mean vectors and covariance matrices with algorithm provided Appendix C. The synthetic data case studies address the following questions:

- Q1:** How sensitive are the investigated methodologies to the initialisation step of the algorithms? Does the dimensionality of the data, the number of observations, a ration of missing values and proportion of outliers have an impact on the stability of the algorithms?
- Q2:** How successful are the newly derived methodologies in estimating the components of model (1) under different proportions of missing values and other perturbation to the data sample? Does the type of the perturbation (row-wise or element-wise) have an impact on the performance of the methodologies?

Hence, the first subsection of Appendix E discuss the stability of investigated the EM algorithms and their ability to estimate the parameters of the model (1) when the assumptions of the PPCA model agree with the distribution of the underlying data. Hence, we test Gaussian PPCA and robust Gaussian PPCA algorithms given the samples from a multivariate Gaussian distribution, and we examine the t-Student frameworks given the data which follows multivariate t-Student distributions discussed in Section 3. We study the performance of the algorithms for the different dimensionality of a single observation and the sample size as well as various proportions of missing values in the data set without corrupting the data sample with any additional noise.

Next, we focus on the efficiency of the algorithms in estimating the parameters of model (1) given the sample data which is corrupted. We generate the data from a multivariate Gaussian distribution which then is perturbed row-wise (the whole observation is perturbed) or element-wise (only single elements of an observation vector are perturbed) by multivariate Laplace distribution with zero mean and the covariance matrix of Gaussian data. We study the performance of the algorithm under different dimensionality of the data, the number of observations, proportion of missing values and proportions of the corrupted sample.

We test the algorithms given 100 simulations and consider various input parameters: sample size $N = 200, 500$; the dimensionality of a single observation $d = 5, 10, 20$; the proportion of missing values $missing = 0\%, 20\%, 50\%$ and proportion of the perturbed sample $perturbed = 20\%, 50\%$. We examine the robust Gaussian PPCA framework for breakdown points $bp = 0.1, 0.2, 0.3, 0.4, 0.45$. We choose the grid of degrees of freedom for t-Student PPCA frameworks to be flexible and reflect both heavy tail assumptions and less robust frameworks, that is $dof = \{ \{1, \dots, 10\}, 15, 20, 30, 50, 75, 100 \}$. In addition, we examine the effect of the robust initialisation on the performance of the algorithms with corresponding input variable $robustInit = TRUE, FALSE$.

E.1. Testing PPCA Algorithms For Non-Perturbed Sample

We examine the recovery of the following static parameters: the mean vector $\mu_{1 \times d}$ and the $k = 3$ first eigenvectors of the covariance matrix $\mathbf{C}_{d \times d} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}_d$ for $d \times k$ real matrix \mathbf{W} . Additionally,

we remove some portion of data equal to the parameter *missing* in order to test how the proposed estimation framework handles presence of incomplete observations.

We simulate 100 estimations of the true parameters for each of the investigated PPCA frameworks and various assumptions on the dimensionality of the data, sample size, the proportion of missingness. We set the sample set to be fixed per dimensionality in order to examine the sensitivity of the algorithms to the initialisation step. Therefore, each of the plots in this sections shows the median of results with corresponding 97.5% empirical quantiles. For t-Student PPCA cases, we keep the parameter of degrees of freedom fixed to its true value $\nu = 4$.

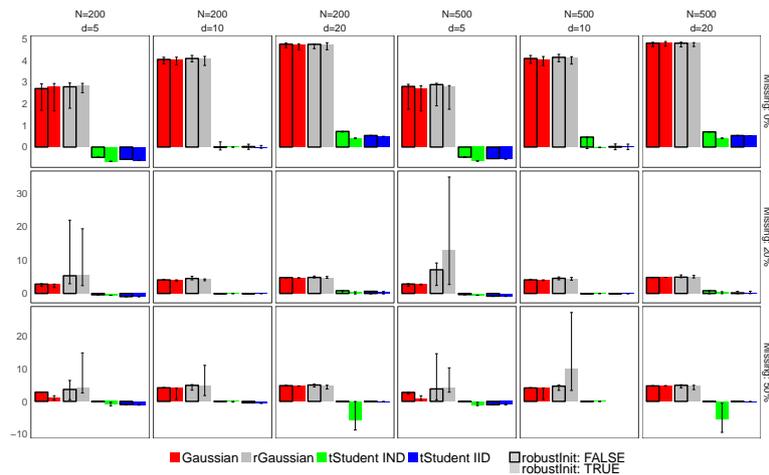


Figure S12. The median of the loglikelihood functions of EM algorithm for: Gaussian PPCA (red), robust Gaussian PPCA (grey), t-Student IND PPCA (green) and t-Student IID PPCA (blue) for (column-wise) different dimensionality of the observation (d) and sample size (N) and (row-wise) proportion of missing values in sample set (*missing*). The bars with solid black borders correspond to the results of the algorithms which have not been initialized with the robust step. The vertical solid black lines highlight the 2.5% and 97.5% empirical quantiles of the obtained results.

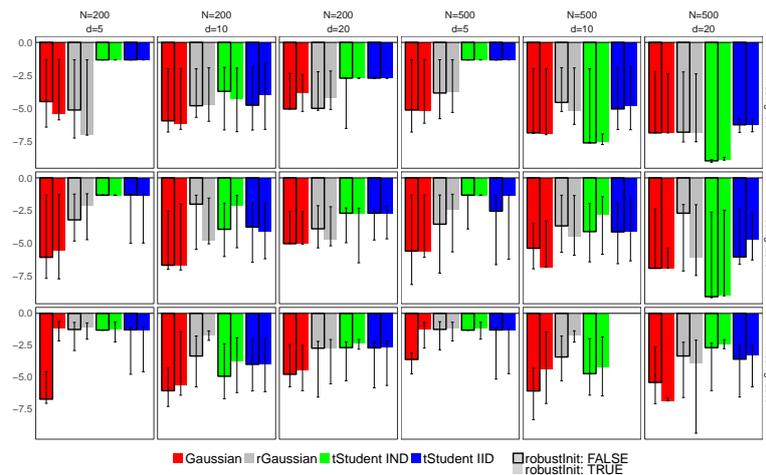


Figure S13. The logarithm of the squared Frobenius norm of the difference between the three first eigen vectors of the true matrix $C = W^T W + \sigma^2 I_d$ and its estimators $\hat{C} = \hat{W}^T \hat{W} + \hat{\sigma}^2 I_d$ given by by EM algorithm for: Gaussian PPCA (red), Robust Gaussian PPCA (grey), t-Student IND PPCA (green) and t-Student IID PPCA (blue) for (column-wise) different dimentionality of a sigle observation (d) and the sample size (N) and (row-wise) the proportion of the missing values in the sample set (*missing*). The bars with solid black borders correspond to the results fo the algorithms which have robust initialisation. The vertical solid black lines highlights the 2.5% and 97.5% quantiles of the results.

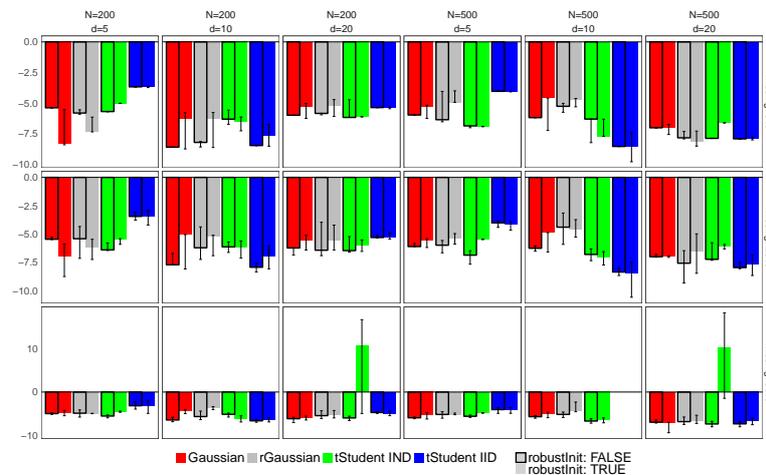


Figure S14. The logarithm of the squared Frobenius norm of the difference between the true value of the vector of means μ and its estimator given by by EM algorithm for: Gaussian PPCA (red), Robust Gaussian PPCA (grey), t-Student IND PPCA (green) and t-Student IID PPCA (blue) for (column-wise) different dimensionality of a single observation (d) and the sample size (N) and (row-wise) the proportion of the missing values in the sample set (*missing*). The bars with solid black borders correspond to the results fo the algorithms which have robust initialisation. The vertical solid black lines highlight the 2.5% and 97.5% quantiles of the results.

Figure S12 illustrates the final loglikelihood functions of the EM algorithms for different variants of PPCA framework. The bars show the median loglikelihood of the corresponding EM algorithms over 100 simulations per case study whereas the vertical solid lines denote 2.5% and 97.5% empirical quantiles of the results. The non-zero interquartile range informs us about the sensitivity of the algorithms to the initialisation step. Recall that we test the algorithms on the datasets which correspond to the PPCA assumptions, hence the magnitude of resulting loglikelihood across different PPCA framework does not inform us which algorithm fits the data most accurately. We can notice that the

algorithms are the most affected by changes to the initialisation when the dimensionality of the data is the lowest regardless of the sample size, especially when the proportion of missing values is the highest. The wide interquartile intervals of robust Gaussian PPCA (grey bars) for $d = 5$ and non zero missingness shows that this methodology is the most sensitive to the initialisation. It is the intuitive outcome since the multivariate frameworks of the robust covariance estimation are very sensitive to their own initialisation what increases the overall sensitivity of the robust Gaussian PPCA framework. On the other hand, the methodology provides solutions with much higher likelihood than the standard Gaussian PPCA for lower dimensionality cases with a presence of missing values

The t-Student PPCA algorithms are less affected by the initialisation step for most of the cases but for the higher dimensionality of the sample data and highest portion of missing values we notice greater sensitivity of t-Student IND PPCA when we do not proceed with robust initialisation. For the other cases related to the two t-Student PPCA frameworks, the robust initialisation neither does decrease nor increase the sensitivity of the algorithm. In addition, robustly initialized t-Student PPCA tends to find solutions in local maximums of likelihoods which have higher values than their equivalents provided by non-robustly initialised algorithms, especially for t-Student IND PPCA.

The convergence of the algorithms to different local maxima of the corresponding loglikelihood function of the EM algorithms results in different estimations of the model's (1) static parameters and therefore the eigenvectors of the covariance matrix differ. The Figure S13 illustrates the discrepancies between the matrix with three most meaningful eigenvectors of the true covariance matrix and its estimation across different cases and methodologies. We use the squared Frobenius norm of the difference between matrices as a measure of the distance.

As expected, the sensitivity of the estimation algorithms to the initialisation step influences the variability of the estimators. The plots in Figure S13 shows the median of the norms with the corresponding 97.2% empirical quantiles. The standard EM algorithm for Gaussian PPCA outperforms its robust equivalent, especially for the incomplete data cases. Even though the robust Gaussian algorithms converged to the higher likelihood of the data, its quality of the recovery of the matrix with eigenvectors is worse than the standard Gaussian PPCA. Recall that the Gaussian PPCA is the least affected methodology by the change of the input parameters: dimensionality, sample sizes and the proportion of missing values.

One may notice the improvement of estimation of the eigenvectors obtained by t-Student frameworks with the increase of the sample size, especially for t-Student IND PPCA. Both methodologies provide the most accurate estimate of the eigenvectors for higher dimensions of the data. The median of the results between the frameworks initialised robustly or not exhibit small dependencies, but the robust initialisation increases their interquartile range, and consequently the certainty.

Figure S14 illustrates the estimation of the mean vector μ . Similarly to the matrix with eigenvectors, we use the squared Frobenius norm of the difference between the true value of the vector of means μ and its estimate to assess the distance between the vectors. The plots show the median of the logarithms of the norms with corresponding 97.5% empirical confidence intervals. The accuracy of the estimation of μ is less volatile in different cases than for the corresponding matrices with eigenvectors. We observe a small decrease of the confidence of the results provided by the algorithms with the robust initialisation, with the exception of the t-Student IND PPCA for the highest proportion of missing values and dimensionality. The algorithms tend to provide poor accuracy of the estimation of μ what explains the low loglikelihood values of the corresponding EM algorithms shown in Figure S12. We can try to improve the results by restraining the convergence criterion.

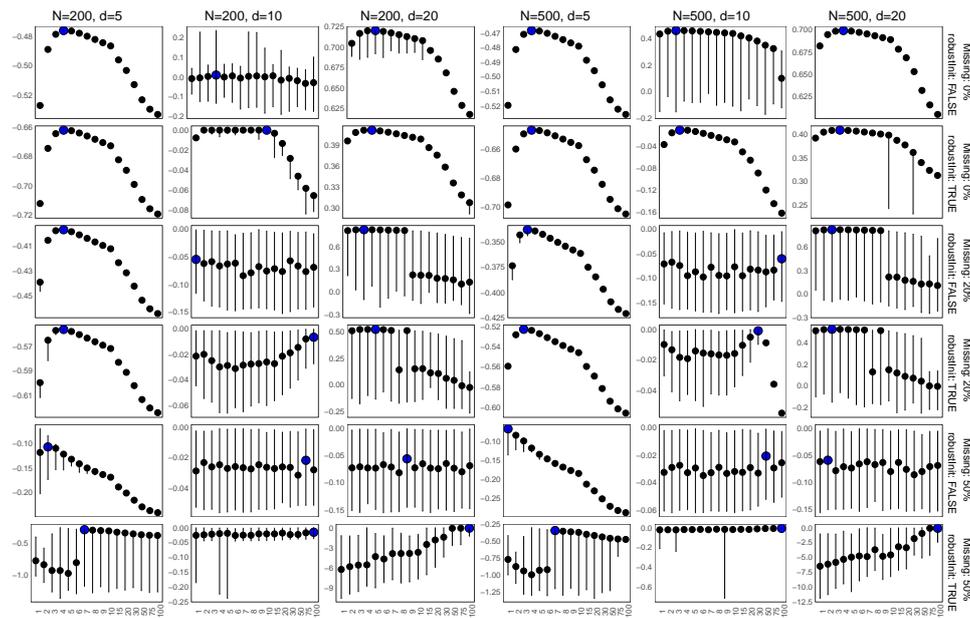


Figure S15. The median of the loglikelihood function of the EM algorithm for t-Student IND PCA across the grid of degrees of freedom (x -axis) over different dimensionality of the observations (d) and the sample size (N)(column-wise) and the proportion of the missing values in the sample set (row-wise). The vertical solid black lines highlight the 2.5% and 97.5% empirical quantiles of the results. The blue colour of the points corresponds to the degree of freedom parameter which is optimal in terms of MLE model selection criterion. The true value of the parameter is equal to 4.

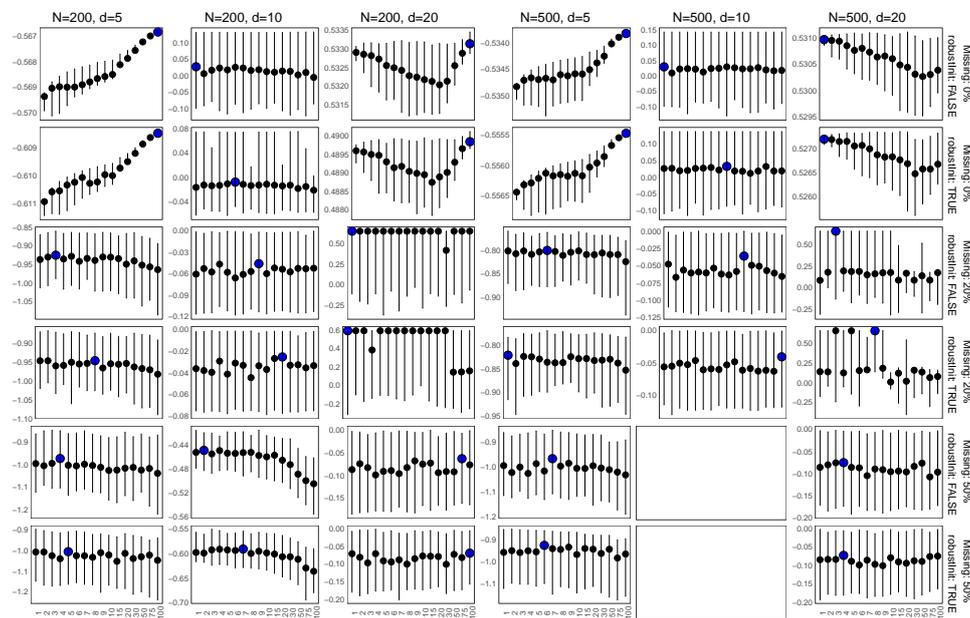


Figure S16. The median of the loglikelihood function of the EM algorithm for t-Student IID PCA across the grid of degrees of freedom (x -axis) over different dimensionality of the observations (d) and the sample size (N)(column-wise) and the proportion of the missing values in the sample set (row-wise). The vertical solid black lines highlight the 2.5% and 97.5% empirical quantiles of the results. The blue colour of the points corresponds to the degree of freedom parameter which is optimal in terms of MLE model selection criterion. The true value of the parameter is equal to 4.

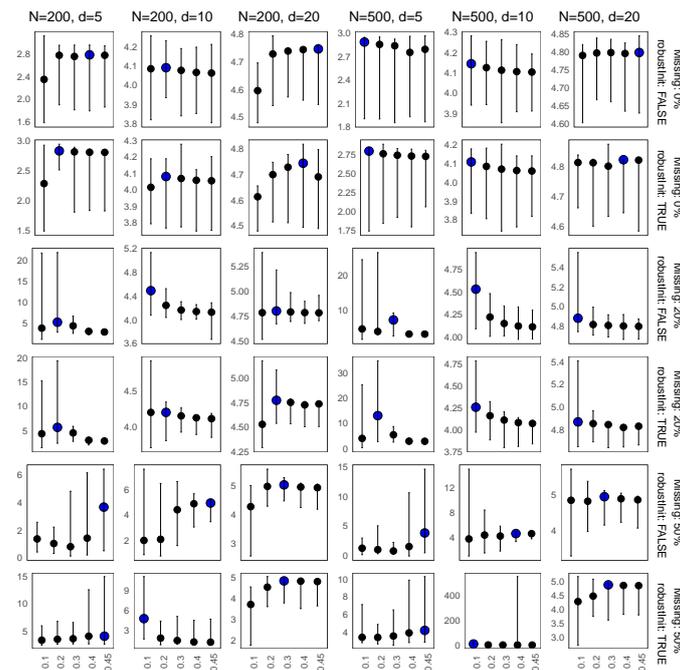


Figure S17. The median logarithm of the likelihood function of EM algorithm for robust Gaussian PPCA across the grid of degrees of freedom (x-axis) over different dimensionality of the observation (d) and sample size (N) and (row-wise) proportion of missing values in sample set ($missing$). The blue colour of bars corresponds to the degree of freedom parameter per case which results in the maximal median value of the corresponding loglikelihood. The bars with red borders correspond to the results for the algorithms which have robust initialisation. The vertical solid black lines highlights the 2.5% and 97.5% empirical quantiles of the results.

We want to verify what is the impact of different assumption on heavy tails of a sample data on the loglikelihood of the t-Student PPCA and robust Gaussian PPCA EM algorithms. Figure S15 and Figure S16 present the median logarithm of the likelihood function per dimensionality, sample size and proportion of missingness across different values of degrees of freedom for t-Student IND and t-Student IID PPCA algorithms whereas Figure S17 illustrates the results for robust Gaussian PPCA across various breakdown points. The maximum median of logarithms is highlighted by blue colour. The row-wise order of the panels corresponds to the different proportions of missing values and the type of the initialisation step

One may notice that the algorithms remind sensitive to the initialisation step regardless of the parameters of degrees of freedom or a breakdown point, especially with the increase of the dimensionality and the ratio of missing value. In addition, the t-Student PPCA algorithms rarely provide the highest values of the loglikelihood functions for their true parameter of degrees of freedom. The corresponding likelihood functions are very flat across considered values of degrees of freedom and therefore the estimation of the parameters is sensitive to numerical error. Consequently, we recommend against the estimation of these parameters via Maximum Loglikelihood Estimation in EM algorithm as it is proceeded with the rest of the linear static parameters of the model (1). Therefore, we use the parameter of degrees of freedom as a model selection criterion and specify its value on a grid.

With regards to the determining the value of the breakdown points in robust Gaussian PPCA, the choice is related to model selection methodology as it is not a component of the model (1) or assumptions on Gaussian PPCA. Figure S17 shows that the optimal breakdown points are not significantly affected by the initialisation methodology. However, we observe that the selection of the values is affected by the dimensionality of the data and the proportion of missing values.

The breakdown point denotes the allowed maximal proportion of a sample corrupted with outlying data which can be handled before by a covariance estimator before giving an incorrect

result. The plots in Figure S17 further confirm that the robust Gaussian PPCA methodology is more sensitive to initialisation step with an increase of missing values, especially for lower breakdown points. The profile loglikelihood of the corresponding EM algorithms is not flat and provides us with clear information which value of the parameter to select, however, the interquartile range of the results tends to be wide especially for the higher proportion of missing values and small dimensionality.

E.2. Testing Robustness of PPCA Algorithms For the Perturbed Sample

In order to investigate the efficiency of the PPCA frameworks in a presence of perturbation to the sample data, we generate the observation data which follows multivariate Gaussian distribution and then corrupt some portion of the sample corresponding to the parameter *perturbed* with the realisation of zero mean multivariate Laplace random vector with covariance matrix \mathbf{C} and the dimensionality which is consistent with the observation vector.

We perturb the observation set in two ways: (1) row-wise when we draw the whole observations and perturb it with the corresponding realisation of the Laplace random vector; or (2) we draw single elements from the sample set which we perturb with corresponding elements from a realisation of the Laplace random vector. The studies aim to identify which methodology is the most robust in the presence of two different perturbation patterns. We examine the methodology under different input parameters: length of the sample, the dimensionality of an observation, the proportion of missing values and the proportion of the corrupted sample.

The studies show that in the presence of the element-wise perturbation, the best methodology which estimates the mean vector μ is t-Student IND regardless of the input parameters. The panels in Figure S18 show the accuracy of the estimations of the three most meaningful eigenvectors of the covariance matrix \mathbf{C} for two patterns of the perturbation: row-wise in the top panel and element-wise in the bottom panel. They illustrate the logarithm of the median of discrepancies between the estimation and the true value of the parameter across different cases and methodologies. We use the squared Frobenius norm of the difference between matrices as a measure of the distance. We remark that the Gaussian PPCA is comparably accurate in the estimation of the matrix with eigenvectors that the robust frameworks when the dimensionality of the data is higher. The robust Gaussian PPCA provides the poorest recovery. The t-Student IND PPCA obtains the best results for the lower dimensionality of the data when the sample is perturbed element-wise or for higher dimensionality when the sample is perturbed row-wise. Also, the high proportion of missing values decrease the performance of the algorithms. With regards to the t-Student IID PPCA, the methodology is most efficient in high dimensional and obtains the best performance among examined frameworks when the data is corrupted element-wise.

The estimation of the vector of means, μ is more consistent across the methodologies and perturbations patters. The panels in Figure display the accuracy of the estimations for row-wise and element-wise perturbation of the sample set, in the top and bottom panel, respectively. We notice, that similarly to the non-perturbed case, the t-Student IND PPCA tends to provide the poor estimation of the parameter when the dimensionality of the data and the proportion of missing values are high. In addition, we observe that in a presence of the row-wise perturbation and high dimensionality of data, the t-Student IID PPCA is the most accurate methodology.

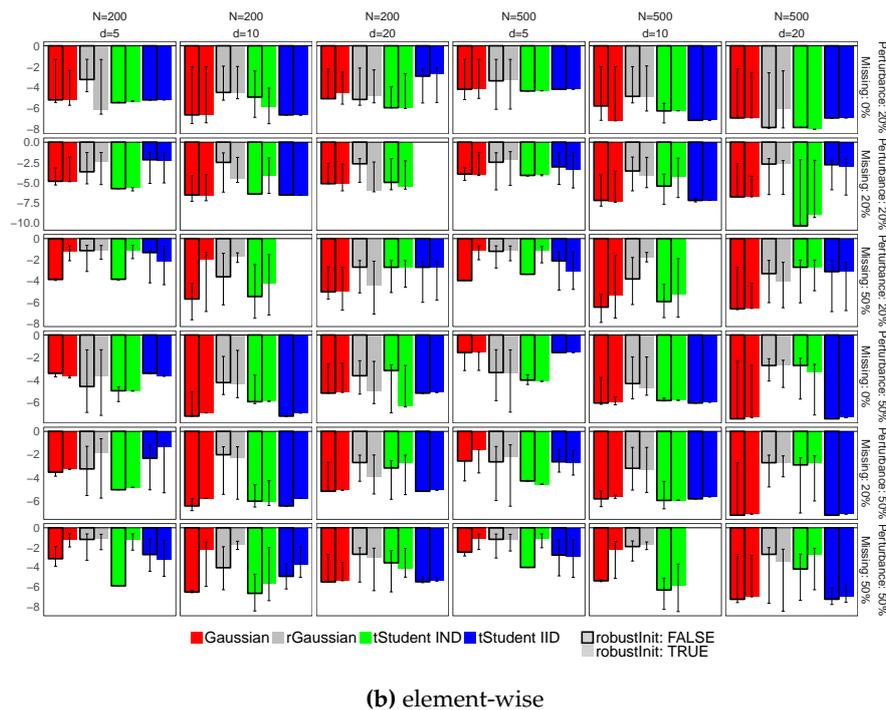
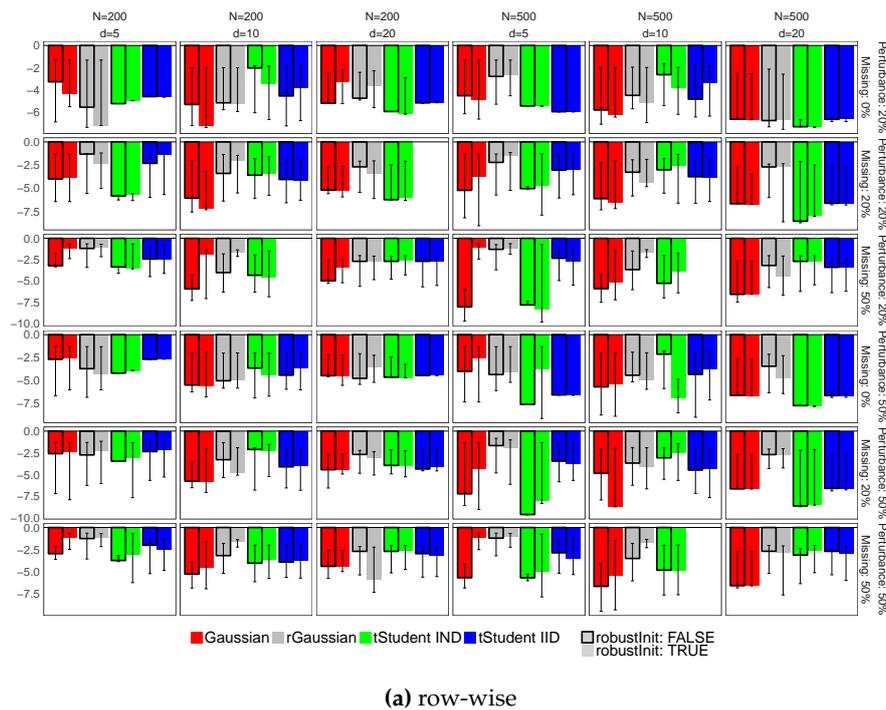
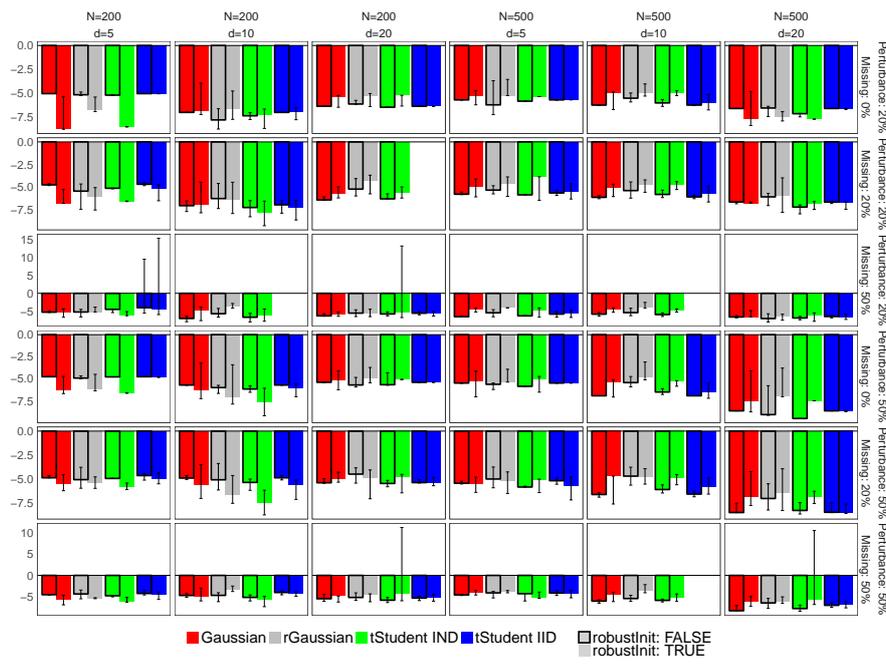
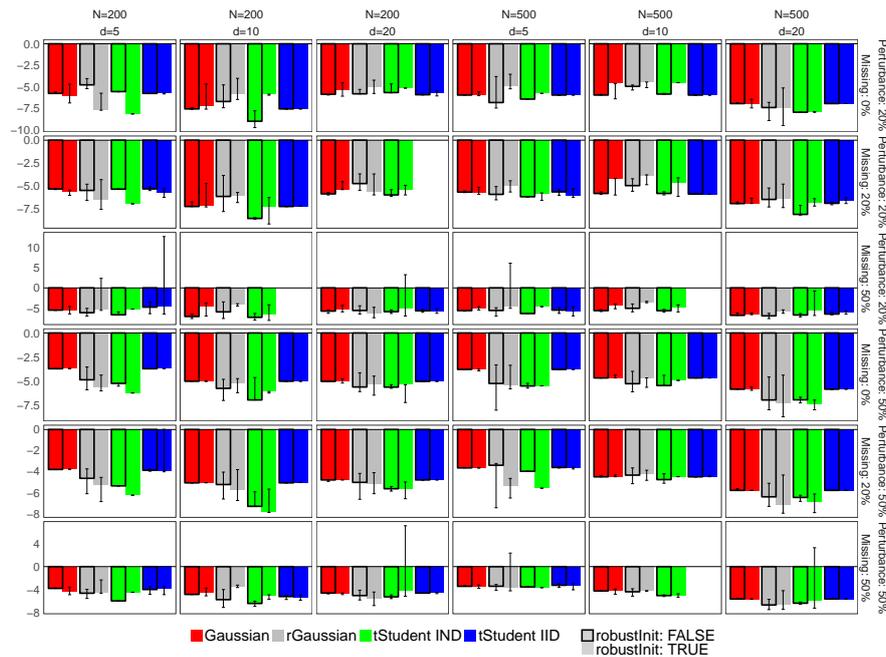


Figure S18. The logarithm of the squared Frobenius norm of the difference between the three first eigen vectors of the true matrix $\mathbf{C} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}_d$ and its estimate $\hat{\mathbf{C}} = \hat{\mathbf{W}}^T \hat{\mathbf{W}} + \sigma^2 \mathbf{I}_d$ given by the following EM algorithms: Gaussian PPCA (red), Robust Gaussian PPCA (grey), t-Student IND PPCA (green) and t-Student IID PPCA (blue) for (column-wise) different dimensionality of a single observation (d) and the sample size (N) and (row-wise) the proportion of the sample which is perturbed with the Laplace noise and the proportion of the missing values in the sample set. The bars with solid black borders correspond to the results for the algorithms which have robust initialisation. The vertical solid black lines highlights the 2.5% and 97.5% quantiles of the results. The top and bottom panels correspond to the row-wise and element-wise perturbation patterns, respectively.



(a) row-wise



(b) element-wise

Figure S19. The logarithm of the squared Frobenius norm of the difference between the true vector of means μ and its estimator given by the following EM algorithms: Gaussian PPCA (red), Robust Gaussian PPCA (grey), t-Student IND PPCA (green) and t-Student IID PPCA (blue) for (column-wise) different dimensionality of a single observation (d) and the sample size (N) and (row-wise) the proportion of the sample which is perturbed with the Laplace noise and the proportion of the missing values in the sample set. The bars with solid black borders correspond to the results of the algorithms which have robust initialization. The vertical solid black lines highlight the 2.5% and 97.5% quantiles of the results. The top and bottom panels correspond to the **row-wise** and **element-wise perturbation patterns**, respectively.

The selected breakdown points for robust Gaussian PPCA are illustrated in Figure S20 for row-wise and element-wise pattern of perturbation in left and right panel, respectively. The proportion of perturbed sample is indicated in the x-axis of the plots. The blue color of bars corresponds to the selected parameters when the algorithm is initialized in a robust way. The optimal values are mostly consistent within the perturbation patterns. We notice that the discrepancies between the values of parameters provided by the robust or non-robust initialisation increase when the proportion of missing values is higher and dimensionality of data is low, especially in the element-wise case. What is more, we observe that the increase of the selected breakdown points follows the increase of the missingness. It is not a surprising outcome since the missing values are projected using the conditional means. The robust covariance estimation algorithms detects more variability and outgoingness in the smoothed data set with filled missing entries and hence requires higher breakdown point which specifies the percentage of sample which is drawn to calculate the sample estimators of a covariance matrix.

With regards to the other robust PPCA methodologies, Figure S21 and Figure S22 shows corresponding optimal degrees of freedom selected for t-Student IND PPCA and t-Student IID PPCA, respectively. The selected values of the parameter are less consistent across different patterns of perturbation as well as dimensionality of the data, the sample size and proportion of missing values than the selected breakdown points. It is a consequence of the flat profile of loglikelihood of the corresponding EM algorithms and therefore their numerical sensitivity. There is small consistency of the results with the proportion of the perturbation, especially for higher dimensionality of the data. However, we remark that t-Student IID PPCA selects higher values of degrees of freedom for element-wise perturbation when the proportion of missing values is low whereas lower degrees of freedom when the missingness is the highest. The selected of degrees of freedom are in more agreement for the row-wise perturbation pattern.

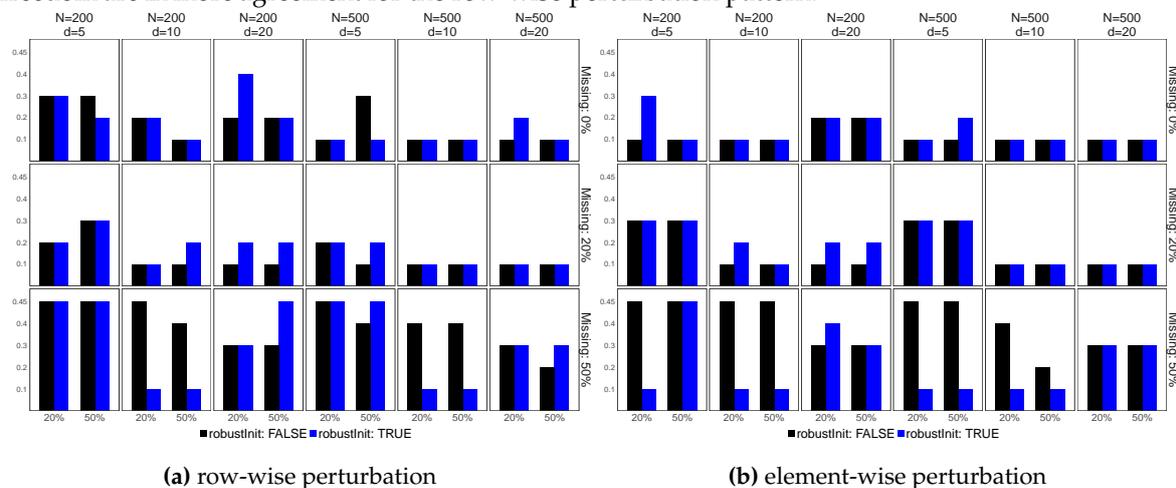


Figure S20. The optimal breakdown points for **robust Gaussian PPCA** (y-axis) across different proportions of perturbed samples with the Laplace noise (x-axis) over different dimensionality of the observation (d) and sample size (N) and (row-wise) the proportion of the missing values in the sample set. The blue color of bars corresponds to the algorithms with the robust initialisation. The left and right panels correspond to the results for row-wise and element-wise perturbations, respectively.

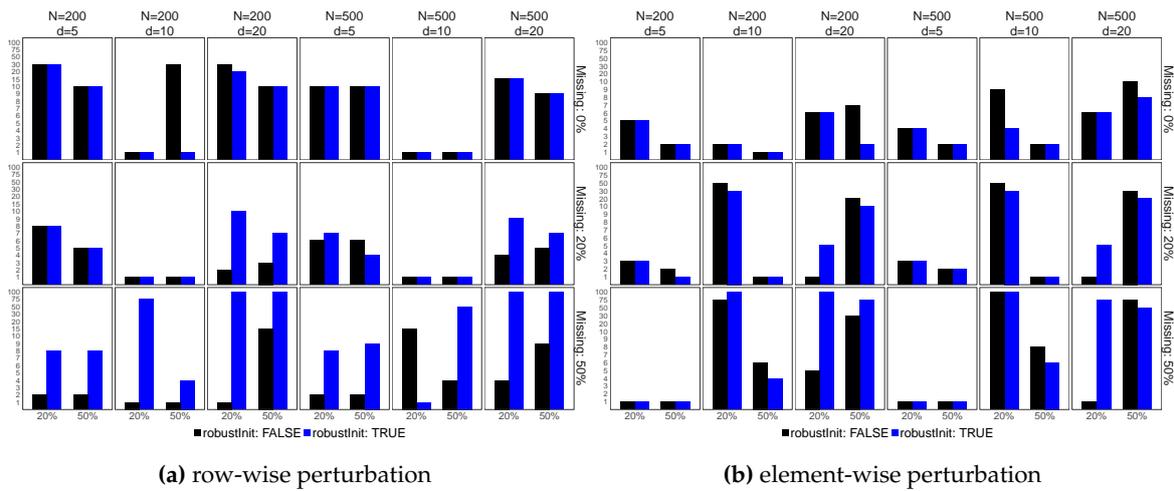


Figure S21. The optimal degrees of freedom for **t-Student IND PPCA** (y-axis) across different proportions of perturbed samples with the Laplace noise (x-axis) over different dimensionality of the observation (d) and sample size (N) and (row-wise) the proportion of the missing values in the sample set. The blue color of bars corresponds to the algorithms with the robust initialization. The left and right panels correspond to the results for row-wise and element-wise perturbations, respectively

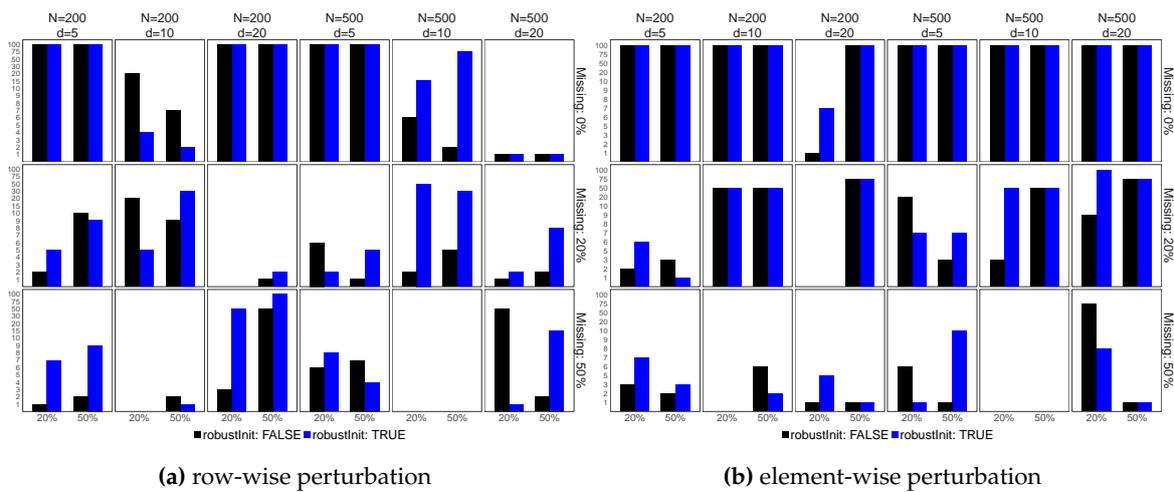


Figure S22. The optimal degrees of freedom for **t-Student IID PPCA** (y-axis) across different proportions of perturbed samples with the Laplace noise (x-axis) over different dimensionality of the observation (d) and sample size (N) and (row-wise) the proportion of the missing values in the sample set. The blue color of bars corresponds to the algorithms with the robust initialization. The left and right panels correspond to the results for row-wise and element-wise perturbations, respectively.

Appendix F. Model Fit Statistics

In the following section we provide the supplementary materials to the result of in-sample and out-of-sample analysis from Section 7.

Table S5. The comparison of the optimal models selected using two employed strategies - by minimum AIC (minAIC) and the lowest complexity (diagAll) for the Euro Libor yield with **1Y swap rate**. The first 22 correspond to the description of the yearly (the first column) in-sample and out-of-sample fit results for the Extended Nelson-Siegel model and the latter to the standard Nelson-Siegel model. The second column, **Crit.**, indicates which selection strategy has been used to decide the assumptions about the static parameters of the models. The columns 3-6 provide the information about the chosen assumptions: **Case** - gives information which matrices are assumed to be diagonal according to the cases in Section 5.4; **Q**- provides information which type of covariance matrix for the observation error term is chosen, heterogeneous (hete) or homogeneous (homo); **R**- provides information which type of covariance matrix for the state equation error term is chosen, heterogeneous (hete), homogeneous (homo) or homogeneous per group (homoP); $\log \lambda$ - provides the value of the shape parameters of the Nelson-Siegel model. The columns 7-10 present the yearly in-sample model fit statistic: **AIC** - Akaike Information Criterion of the models, **BIC** - Bayesian Information Criterion of the models, **MSE** - the logarithm of the mean square errors of the in-sample predictions $y_{t|t-1}$ obtained by the Kalman Filter, **L** - the logarithm of the Kalman Filter likelihood. The columns 11-12 show the forecasting performance of the models measured by the mean square error of prediction (MSEP) being an averaged squared residuals between the out-of-sample prediction provided by the Kalman Filter, $y_{t|t-1}$ and observed values: **1D** - the logarithm of 1 day MSEP, **1W** - the logarithm of 1 week MSEP, **1M** - the logarithm of 1 month MSEP. The last two columns refer to the information about the optimal factors chosen for Extended Nelson-Siegel model: **PPCA** - the PPCA methodology used to obtain the factors from data sets, 1- Gaussian, 2- t-Student IND, 3 - t-Student IID; **Model** - the acronyms of the examined Extended Nelson-Siegel models listed in Section 7.1

| | Year | Crit. | Static Params | | | | In-sample | | | | Out-of-sample | | | Factors | |
|------------------------------|------|---------|---------------|------|-------|----------------|-----------|-----------|-------|----------|---------------|-------|-------|---------|-------|
| | | | Case | Q | R | $\log \lambda$ | AIC | BIC | MSE | L | 1D | 1W | 1M | PPCA | Model |
| Extended Nelson-Siegel model | 2006 | diagAll | 1 | hete | hete | -1.42 | -16200.80 | -16197.25 | -4.36 | 8101.40 | -6.20 | -4.96 | -3.59 | 2 | M2 |
| | | minAIC | 2 | hete | hete | -2.31 | -17898.52 | -17894.97 | -5.10 | 8950.26 | -6.35 | -4.77 | -2.74 | 1 | M2 |
| | 2007 | diagAll | 1 | hete | hete | -0.96 | -14227.19 | -14223.63 | -3.99 | 7114.59 | -4.92 | -3.42 | -1.94 | 1 | M2 |
| | | minAIC* | 4 | homo | hete | -0.40 | -8220.76 | -8217.20 | -6.12 | 4111.38 | -4.39 | -3.27 | -1.75 | 3 | M2 |
| | 2008 | diagAll | 1 | hete | hete | -0.48 | -10153.47 | -10149.90 | -3.81 | 5077.73 | -5.46 | -3.56 | -2.05 | 3 | M3 |
| | | minAIC | 1 | hete | hete | -0.48 | -10153.47 | -10149.90 | -3.81 | 5077.73 | -5.46 | -3.56 | -2.05 | 3 | M3 |
| | 2009 | diagAll | 1 | hete | hete | -0.48 | -14214.22 | -14210.65 | -5.77 | 7108.11 | -5.66 | -3.16 | -1.53 | 2 | M2 |
| | | minAIC* | 1 | homo | hete | -0.40 | -11494.12 | -11490.55 | -5.73 | 5748.06 | -5.19 | -2.47 | -0.66 | 2 | M2 |
| | 2010 | diagAll | 1 | hete | hete | -0.64 | -14898.48 | -14894.92 | -5.81 | 7450.24 | -6.02 | -4.51 | -2.68 | 2 | M2 |
| | | minAIC* | 1 | homo | homoP | -2.31 | -11994.02 | -11990.45 | -5.84 | 5998.01 | -5.04 | -4.08 | -2.52 | 2 | M2 |
| | 2011 | diagAll | 1 | hete | hete | -0.56 | -13025.47 | -13021.91 | -5.07 | 6513.74 | -6.02 | -3.98 | -1.70 | 1 | M3 |
| | | minAIC | 1 | homo | homo | -0.48 | -10689.50 | -10685.94 | -5.61 | 5345.75 | -5.95 | -5.18 | -4.16 | 1 | M2 |
| | 2012 | diagAll | 1 | hete | hete | -0.74 | -15068.73 | -15065.16 | -5.76 | 7535.36 | -7.26 | -5.82 | -4.17 | 2 | M2 |
| | | minAIC | 1 | homo | hete | -0.64 | -11603.07 | -11599.51 | -6.31 | 5802.54 | -6.96 | -5.79 | -4.23 | 1 | M2 |
| | 2013 | diagAll | 1 | hete | hete | -0.74 | -16255.27 | -16251.71 | -6.72 | 8128.64 | -7.19 | -5.88 | -3.74 | 3 | M2 |
| | | minAIC | 1 | hete | hete | -0.74 | -16255.27 | -16251.71 | -6.72 | 8128.64 | -7.19 | -5.88 | -3.74 | 3 | M2 |
| | 2014 | diagAll | 1 | hete | hete | -0.96 | -17049.46 | -17045.90 | -6.47 | 8525.73 | -7.01 | -5.21 | -3.01 | 2 | M2 |
| | | minAIC | 2 | hete | homoP | -0.84 | -18140.85 | -18137.29 | -6.41 | 9071.43 | -7.16 | -5.48 | -3.32 | 2 | M2 |
| | 2015 | diagAll | 1 | hete | hete | -0.96 | -17853.02 | -17849.45 | -6.67 | 8927.51 | -7.69 | -6.29 | -4.38 | 1 | M2 |
| | | minAIC | 4 | hete | hete | -0.96 | -20499.40 | -20495.83 | -6.71 | 10250.70 | -7.26 | -5.15 | -2.97 | 1 | M2 |
| | 2016 | diagAll | 1 | hete | hete | -1.09 | -17814.46 | -17810.90 | -7.15 | 8908.23 | | | | 1 | M2 |
| | | minAIC | 1 | homo | hete | -1.09 | -15528.66 | -15525.09 | -6.41 | 7765.33 | | | | 1 | M2 |
| Nelson-Siegel model | 2006 | diagAll | 1 | hete | hete | -1.42 | -12327.83 | -12324.27 | -4.07 | 6164.91 | -4.01 | -3.85 | -3.26 | | |
| | | minAIC | 1 | hete | hete | -1.42 | -12327.83 | -12324.27 | -4.07 | 6164.91 | -4.01 | -3.85 | -3.26 | | |
| | 2007 | diagAll | 1 | hete | hete | -0.96 | -12511.59 | -12508.03 | -3.23 | 6256.79 | -2.24 | -2.36 | -1.75 | | |
| | | minAIC* | 2 | hete | homo | -0.96 | -12709.18 | -12705.63 | -3.37 | 6355.59 | -2.18 | -2.33 | -1.87 | | |
| | 2008 | diagAll | 1 | hete | hete | -0.48 | -7160.47 | -7156.90 | -2.64 | 3581.24 | -4.43 | -3.89 | -2.65 | | |
| | | minAIC | 1 | hete | hete | -0.48 | -7160.47 | -7156.90 | -2.64 | 3581.24 | -4.43 | -3.89 | -2.65 | | |
| | 2009 | diagAll | 1 | hete | hete | -0.48 | -11357.15 | -11353.59 | -4.47 | 5679.58 | -3.80 | -2.76 | -1.38 | | |
| | | minAIC* | 2 | hete | hete | -0.56 | -11664.50 | -11660.93 | -4.49 | 5833.25 | -4.60 | -3.14 | -1.51 | | |
| | 2010 | diagAll | 1 | hete | hete | -0.64 | -12052.63 | -12049.06 | -4.31 | 6027.31 | -4.18 | -3.62 | -2.34 | | |
| | | minAIC* | 2 | hete | hete | -2.31 | -12145.68 | -12142.11 | -4.95 | 6073.84 | -3.22 | -2.95 | -1.91 | | |
| | 2011 | diagAll | 1 | hete | hete | -0.56 | -10380.08 | -10376.52 | -4.26 | 5191.04 | -3.49 | -3.18 | -2.12 | | |
| | | minAIC* | 2 | hete | hete | -0.56 | -10754.53 | -10750.97 | -4.29 | 5378.26 | -3.46 | -3.23 | -2.39 | | |
| | 2012 | diagAll | 1 | hete | homo | -0.74 | -9508.23 | -9504.67 | -4.21 | 4755.12 | -4.03 | -3.97 | -3.03 | | |
| | | minAIC | 3 | hete | homo | -0.74 | -9518.42 | -9514.86 | -4.35 | 4760.21 | -4.47 | -3.86 | -1.42 | | |
| | 2013 | diagAll | 1 | hete | homo | -0.96 | -10788.14 | -10784.57 | -5.13 | 5395.07 | -5.34 | -4.87 | -3.10 | | |
| | | minAIC | 2 | hete | hete | -0.96 | -11464.71 | -11461.15 | -5.37 | 5733.36 | -5.44 | -4.85 | -2.82 | | |
| | 2014 | diagAll | 1 | hete | hete | -0.96 | -11274.46 | -11270.90 | -5.40 | 5638.23 | -5.83 | -4.93 | -2.92 | | |
| | | minAIC | 3 | hete | hete | -0.96 | -11378.09 | -11374.53 | -5.37 | 5690.04 | -5.73 | -4.88 | -3.33 | | |
| | 2015 | diagAll | 1 | hete | homo | -0.96 | -13368.97 | -13365.40 | -5.85 | 6685.48 | -5.46 | -5.09 | -3.98 | | |
| | | minAIC | 4 | hete | homo | -0.96 | -14395.09 | -14391.52 | -6.06 | 7198.54 | -5.65 | -3.95 | -1.54 | | |
| | 2016 | diagAll | 1 | hete | hete | -1.09 | -12214.66 | -12211.10 | -5.54 | 6108.33 | | | | | |
| | | minAIC | 2 | hete | homo | -1.09 | -12990.60 | -12987.04 | -5.85 | 6496.30 | | | | | |

Table S6. The comparison of the optimal models selected using two employed strategies - by minimum AIC (minAIC) and the lowest complexity (diagAll) for the Euro Libor yield with **1Y ICE rate**. The first 22 correspond to the description of the yearly (the first column) in-sample and out-of-sample fit results for the Extended Nelson-Siegel model and the latter to the standard Nelson-Siegel model. The second column, **Crit.**, indicates which selection strategy has been used to decide the assumptions about the static parameters of the models. The columns 3-6 provide the information about the chosen assumptions: **Case** - gives information which matrices are assumed to be diagonal according to the cases in Section 5.4; **Q**- provides information which type of covariance matrix for the observation error term is chosen, heterogeneous (hete) or homogeneous (homo); **R**- provides information which type of covariance matrix for the state equation error term is chosen, heterogeneous (hete), homogeneous (homo) or homogeneous per group (homoP); **log λ** - provides the value of the shape parameters of the Nelson-Siegel model. The columns 7-10 present the yearly in-sample model fit statistic: **AIC** - Akaike Information Criterion of the models, **BIC** - Bayesian Information Criterion of the models, **MSE** - the logarithm of the mean square errors of the in-sample predictions $y_{t|t-1}$ obtained by the Kalman Filter, **L** - the logarithm of the Kalman Filter likelihood. The columns 11-12 show the forecasting performance of the models measured by the mean square error of prediction (MSEP) being an averaged squared residuals between the out-of-sample prediction provided by the Kalman Filter, $y_{t|t-1}$ and observed values: **1D** - the logarithm of 1 day MSEP, **1W** - the logarithm of 1 week MSEP, **1M** - the logarithm of 1 month MSEP. The last two columns refer to the information about the optimal factors chosen for Extended Nelson-Siegel model: **PPCA** - the PPCA methodology used to obtain the factors from data sets, 1- Gaussian, 2- t-Student IND, 3 - t-Student IID; **Model** - the acronyms of the examined Extended Nelson-Siegel models listed in Section 7.1.

| Year | Crit. | Static Params | | | | In-sample | | | | Out-of-sample | | | Factors | | |
|------------------------------|---------|---------------|------|--------------|-----------|-----------|-----------|-----------|----------|---------------|-------|-------|---------|-------|----|
| | | Case | Q | R | log λ | AIC | BIC | MSE | L | 1D | 1W | 1M | PPCA | Model | |
| Extended Nelson-Siegel model | 2006 | diagAll | 1 | hete | hete | -1.42 | -15794.25 | -15790.70 | -4.30 | 7898.12 | -6.35 | -4.93 | -3.41 | 2 | M2 |
| | minAIC | 1 | hete | hete | -1.42 | -15794.25 | -15790.70 | -4.30 | 7898.12 | -6.35 | -4.93 | -3.41 | 2 | M2 | |
| | 2007 | diagAll | 1 | hete | hete | -0.76 | -14609.24 | -14605.70 | -3.67 | 7305.62 | -4.90 | -3.33 | -2.01 | 1 | M2 |
| | minAIC | 2 | hete | homoP | -0.83 | -14127.90 | -14124.35 | -4.43 | 7064.95 | -5.00 | -3.49 | -2.08 | 2 | M2 | |
| | 2008 | diagAll* | 1 | hete | hete | -0.52 | -9680.15 | -9676.59 | -3.98 | 4841.08 | -5.70 | -3.96 | -2.57 | 3 | M3 |
| | minAIC | 2 | hete | hete | -0.52 | -13679.93 | -13676.37 | -3.93 | 6840.96 | -5.41 | -3.47 | -1.59 | 3 | M3 | |
| | 2009 | diagAll | 1 | hete | hete | -0.52 | -13527.22 | -13523.65 | -5.64 | 6764.61 | -5.60 | -3.07 | -1.45 | 2 | M2 |
| | minAIC | 4 | hete | hete | -0.52 | -16645.06 | -16641.49 | -4.81 | 8323.53 | -5.32 | -3.17 | -1.87 | 3 | M3 | |
| | 2010 | diagAll | 1 | hete | hete | -2.18 | -15298.46 | -15294.90 | -6.20 | 7650.23 | -5.52 | -3.69 | -2.22 | 2 | M2 |
| | minAIC | 4 | hete | hete | -2.51 | -18306.79 | -18303.23 | -6.33 | 9154.40 | -5.70 | -4.65 | -2.88 | 2 | M2 | |
| | 2011 | diagAll | 1 | hete | hete | -0.52 | -13003.63 | -13000.07 | -5.19 | 6502.81 | -6.16 | -4.23 | -1.99 | 1 | M3 |
| | minAIC | 2 | hete | hete | -0.52 | -16224.58 | -16221.02 | -4.99 | 8113.29 | -5.93 | -4.01 | -1.79 | 1 | M3 | |
| | 2012 | diagAll | 1 | hete | hete | -0.52 | -15544.17 | -15540.61 | -6.23 | 7773.09 | -7.31 | -5.79 | -4.13 | 2 | M2 |
| | minAIC | 1 | hete | hete | -0.52 | -15544.17 | -15540.61 | -6.23 | 7773.09 | -7.31 | -5.79 | -4.13 | 2 | M2 | |
| | 2013 | diagAll | 1 | hete | hete | -0.83 | -15706.06 | -15702.49 | -6.55 | 7854.03 | -7.18 | -5.74 | -3.73 | 3 | M2 |
| | minAIC | 2 | hete | hete | -0.83 | -18964.99 | -18961.43 | -6.36 | 9483.50 | -7.39 | -5.61 | -3.38 | 1 | M3 | |
| | 2014 | diagAll | 1 | hete | hete | -0.91 | -16922.96 | -16919.39 | -6.42 | 8462.48 | -7.05 | -5.25 | -3.17 | 2 | M2 |
| | minAIC | 2 | hete | hete | -0.91 | -19222.55 | -19218.98 | -6.60 | 9612.27 | -7.31 | -6.14 | -4.43 | 1 | M3 | |
| | 2015 | diagAll | 1 | hete | hete | -0.99 | -17770.05 | -17766.48 | -6.63 | 8886.02 | -7.77 | -6.31 | -4.49 | 1 | M2 |
| | minAIC | 2 | hete | hete | -0.99 | -20003.81 | -20000.24 | -6.63 | 10002.90 | -7.75 | -6.29 | -4.47 | 2 | M2 | |
| | 2016 | diagAll | 1 | hete | hete | -0.99 | -17189.57 | -17186.00 | -6.98 | 8595.78 | -7.75 | -6.29 | -4.47 | 1 | M2 |
| | minAIC | 2 | hete | homoPerGroup | -0.99 | -18389.86 | -18386.30 | -6.93 | 9195.93 | -7.75 | -6.29 | -4.47 | 2 | M2 | |
| Nelson-Siegel model | 2006 | diagAll | 1 | hete | hete | -1.42 | -11843.46 | -11839.91 | -3.99 | 5922.73 | -3.93 | -3.78 | -2.94 | | |
| | minAIC | 1 | hete | hete | -1.42 | -11843.46 | -11839.91 | -3.99 | 5922.73 | -3.93 | -3.78 | -2.94 | | | |
| | 2007 | diagAll | 1 | hete | homo | -0.83 | -12007.74 | -12004.19 | -3.25 | 6004.87 | -2.07 | -2.20 | -1.82 | | |
| | minAIC | 2 | hete | homo | -0.83 | -12426.62 | -12423.07 | -3.29 | 6214.31 | -1.91 | -2.06 | -1.86 | | | |
| | 2008 | diagAll* | 1 | hete | hete | -0.52 | -6660.50 | -6656.94 | -3.03 | 3331.25 | -4.10 | -3.74 | -2.81 | | |
| | minAIC | 2 | hete | hete | -0.52 | -6967.12 | -6963.56 | -3.09 | 3484.56 | -4.03 | -3.27 | -1.59 | | | |
| | 2009 | diagAll | 1 | hete | hete | -0.52 | -10367.44 | -10363.88 | -4.27 | 5184.72 | -3.77 | -2.77 | -1.26 | | |
| | minAIC | 2 | hete | hete | -0.52 | -10536.63 | -10533.06 | -4.28 | 5269.31 | -3.70 | -2.75 | -1.33 | | | |
| | 2010 | diagAll | 1 | hete | hete | -2.18 | -10842.19 | -10838.62 | -3.66 | 5422.09 | -2.29 | -2.29 | -1.78 | | |
| | minAIC | 4 | hete | hete | -2.51 | -11453.43 | -11449.87 | -3.83 | 5727.72 | -2.44 | -2.37 | -1.87 | | | |
| | 2011 | diagAll | 1 | hete | hete | -0.52 | -8712.66 | -8709.10 | -3.40 | 4357.33 | -2.90 | -2.90 | -2.23 | | |
| | minAIC | 2 | hete | hete | -0.52 | -9677.18 | -9673.62 | -3.39 | 4839.59 | -2.88 | -2.88 | -2.49 | | | |
| | 2012 | diagAll | 1 | hete | hete | -0.52 | -8734.87 | -8731.31 | -2.90 | 4368.44 | -3.86 | -3.69 | -3.03 | | |
| | minAIC | 1 | hete | hete | -0.52 | -8734.87 | -8731.31 | -2.90 | 4368.44 | -3.86 | -3.69 | -3.03 | | | |
| | 2013 | diagAll | 1 | hete | hete | -0.83 | -11474.72 | -11471.15 | -4.69 | 5738.36 | -4.30 | -4.24 | -2.97 | | |
| | minAIC | 1 | hete | hete | -0.83 | -11474.72 | -11471.15 | -4.69 | 5738.36 | -4.30 | -4.24 | -2.97 | | | |
| 2014 | diagAll | 1 | hete | hete | -0.91 | -10448.68 | -10445.12 | -4.43 | 5225.34 | -4.86 | -4.42 | -2.92 | | | |
| minAIC | 2 | hete | hete | -0.91 | -10487.17 | -10483.60 | -4.49 | 5244.58 | -4.91 | -4.48 | -3.05 | | | | |
| 2015 | diagAll | 1 | hete | homo | -0.99 | -12733.24 | -12729.67 | -4.94 | 6367.62 | -4.37 | -4.28 | -3.61 | | | |
| minAIC | 2 | hete | homo | -0.99 | -13518.68 | -13515.11 | -5.08 | 6760.34 | -4.53 | -4.43 | -3.63 | | | | |
| 2016 | diagAll | 1 | hete | hete | -0.99 | -10298.73 | -10295.16 | -4.69 | 5150.36 | | | | | | |
| minAIC | 4 | hete | homo | -0.99 | -11304.45 | -11300.89 | -4.71 | 5653.23 | | | | | | | |

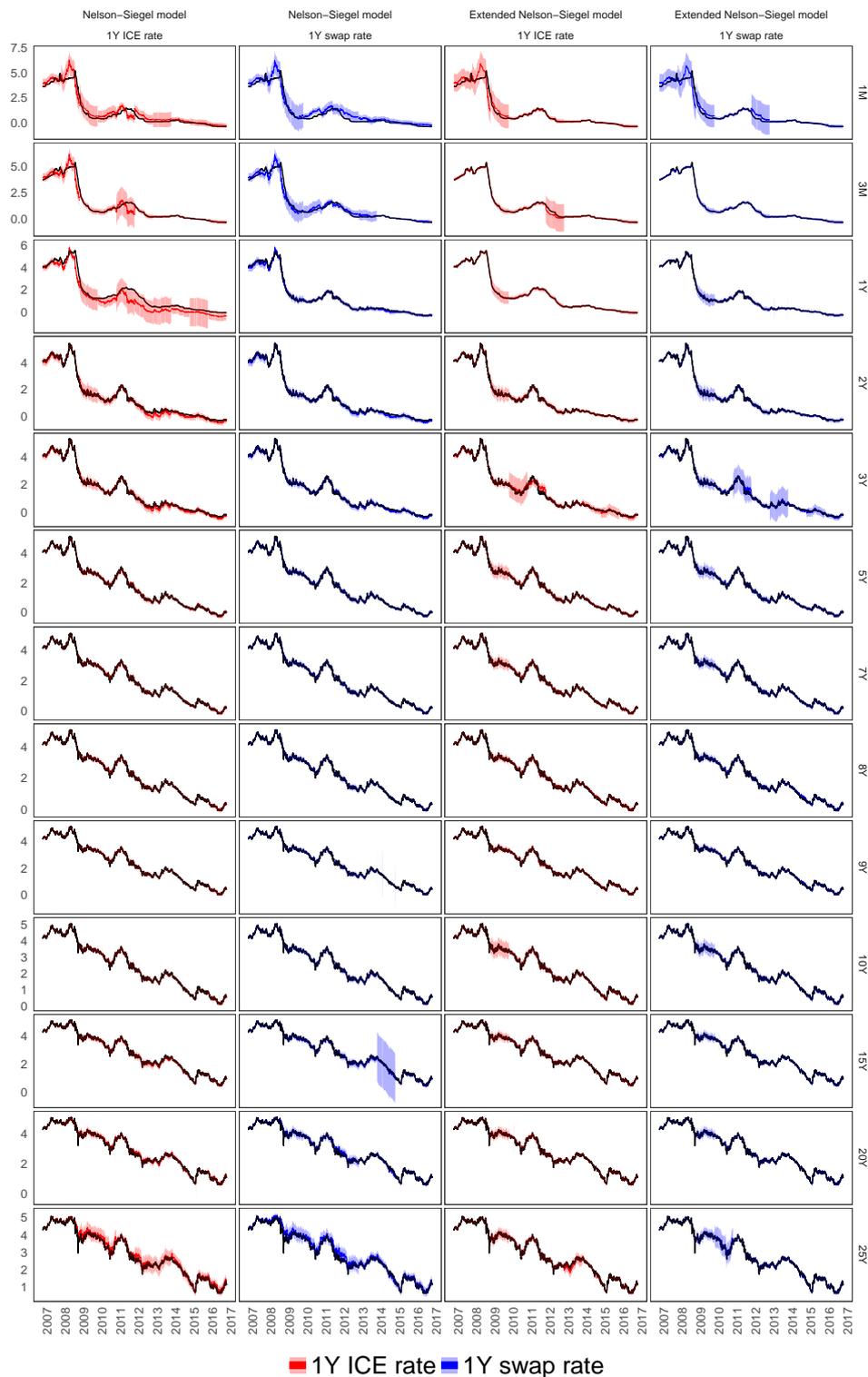


Figure S23. The Euro Libor yield curve (black line) with the one step ahead out-of-sample prediction given by the conditional mean $y_{T+s+1|T+s}$ with corresponding 97.5% confidence intervals over examined period. The column-wise order of panels correspond to different classes of models, the Nelson-Siegel or Extended Nelson-Siegel model, calibration on two data sets with different 1Y rate (colors of lines and shading). The row-wise order corresponds to the evolution of rates at different maturities over time.

References

Gupta, A. K. and D. K. Nagar. 1999. *Matrix Variate Distributions*. Chapman and Hall/CRC.

Tipping, M. E. and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society*. 61(3), 622–661.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).